SPEND WHERE IT COUNTS: DYNAMIC BUDGET ALLOCATION FOR MULTI-TURN LLM EVALUATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Evaluating and predicting the performance of large language models (LLMs) in multi-turn conversational settings is critical yet computationally expensive; key events—e.g., jailbreaks or successful task completion by an agent—often emerge only after repeated interactions. These events might be rare and, under any feasible computational budget, remain unobserved. Recent conformal survival frameworks construct reliable lower predictive bounds (LPBs) on the number of iterations to trigger the event of interest, but rely on static budget allocation that prohibits adaptivity in multi-turn setups. To address this, we introduce *Dynamic Allocation via PRojected Optimization* (DAPRO), a theoretically valid dynamic budget allocation framework for bounding the time-to-event in multi-turn LLM interactions. We prove that DAPRO satisfies the expected budget constraint and provides distribution-free, finite-sample coverage guarantees without requiring the conditional independence between censoring and event times assumed by prior conformal survival approaches. A key theoretical contribution is a novel coverage bound that scales with the square root of the mean censoring weight over only a subset of the samples rather than the worst-case weight, yielding tighter guarantees than prior work. Furthermore, DAPRO can be employed to obtain unbiased estimates of population-level evaluation metrics, such as the jailbreak rate, under limited computing resources. Comprehensive experiments across agentic task success, LLM jailbreaks, toxic content generation, and RAG hallucinations demonstrate that DAPRO achieves coverage closer to the nominal level with lower variance than static baselines, while not exceeding the budget constraint in expectation.

1 INTRODUCTION

AI systems are moving from one-shot prediction to sequential interaction (Park et al., 2023; Wang et al., 2024a). For example, an agent may need many steps to solve a complex task. Evaluating such systems is challenging because important outcomes may emerge only after many interactions, while running these interactions can be computationally expensive. Rather than asking only whether an AI system will eventually succeed or fail at a task, we ask a richer question: how long will it take for an important event to occur, if ever? For example, how many interactions are needed for a jailbreak to succeed or for an agent to complete a task?

Our framework applies broadly to evaluating time-to-event—time for an outcome of interest—in sequential AI systems. For simplicity, we use LLM safety as a running example, where the event of interest may correspond to an unsafe response, such as toxic language (Gehman et al., 2020) or a hallucination (Ji et al., 2023). In agentic utility settings, the event could be a successful task completion, e.g., identifying a bug in code (Jimenez et al., 2024) (see Appendix B). We assume that event occurrence is determined by an audit function, such as an LLM-as-a-judge (Zheng et al., 2023). Since these events can be rare, particularly for well-aligned models, Davidov et al. (2026) recently introduced *time-to-unsafe* to quantify model safety. In our setting, this metric represents the number of conversational interactions required to elicit an unsafe response given an initial prompt.

Using this metric, LLM safety or utility assessment can be framed as a survival analysis task (Candès et al., 2023; Gui et al., 2024; Sesia & Svetnik, 2024; 2025), which has two advantages. First, the metric is well-defined despite the time dependence of LLM outputs. Second, it enables constructing calibrated *lower predictive bounds* (LPBs) $\hat{L}(X_{\text{test}})$ or *upper predictive bounds* (UPBs) $\hat{U}(X_{\text{test}})$ for

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.

the unknown time-to-event T_{test} of a test prompt X_{test}, with formal finite-sample coverage guarantees $\mathbb{P}(T_{\text{test}} \geq \hat{L}(X_{\text{test}})) \geq 1-\alpha$ and $\mathbb{P}(T_{\text{test}} \leq \hat{U}(X_{\text{test}})) \geq 1-\alpha$. Here $1-\alpha$ is the target coverage rate, e.g., $1-\alpha = 90\%$. Informally, an LPB indicates that the event is unlikely to occur before $\hat{L}(X_{\text{test}})$, while a UPB indicates that it is likely to occur by $\hat{U}(X_{\text{test}})$. Thus, LPBs are natural for events such as unsafe responses, whereas UPBs are useful for events such as agentic task completion.

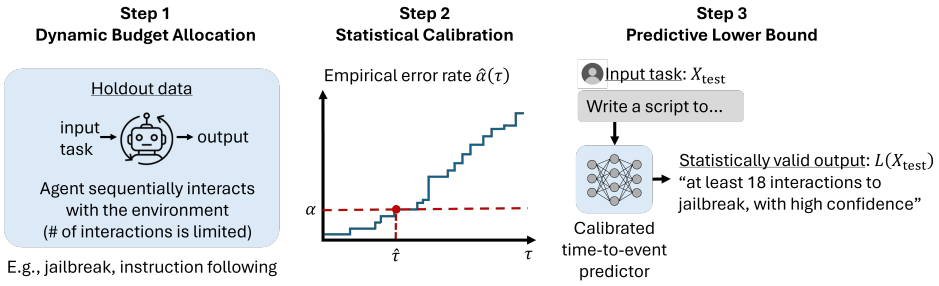

Figure 1: Illustration of our framework: (i) collecting data via dynamic budget allocation; (ii) calibrating a pre-trained model; and (iii) constructing calibrated predictive bounds for time-to-event.

To construct these bounds, we must first acquire a dataset of annotated conversations under the budget constraint. For example, jailbreaking requires running an attacking algorithm starting from a set of initial prompts and continuously evaluating the outputs of the target model using the auditor; similarly, evaluating agentic systems may require running agents over multiple interactions until task completion. In our setup, we suppose the expected number of interactions is constrained by a computational budget $B \in \mathbb{N}$, e.g., $B = 1,000$ interaction rounds in expectation. Formulating the budget constraint in expectation follows the active learning literature (Dasgupta et al., 2007; Altman, 2021; Zrnic & Candès, 2024). To satisfy this budget constraint, we set a censoring time C_i for each sequence i, which limits the maximal trajectory length. Importantly, under such a constraint, the event of interest may not be observed for every sequence.

To provide coverage guarantees for unseen test points, we assume that the calibration and test data are drawn i.i.d. from the same setting; in particular, the prompt distribution and the agent (or attacker), target, and audit models are fixed. This assumption is natural in settings where calibration and test interactions involve different users drawn from the same underlying population, such as educational AI tutors (Xie et al., 2025) or customer-service chatbots (Pandya & Holia, 2023). Even under this assumption, constructing valid bounds from censored data remains challenging: naively calibrating using the partially observed event times can lead to undercoverage, as we observe empirically.

Davidov et al. (2026) propose an optimized budget allocation strategy that aims to minimize the coverage variance of the resulting bounds. Note that obtaining a lower variance is crucial in practice, where only a single dataset is typically available. A lower variance means that our bounds attain a coverage rate closer to the target level $1-\alpha$ given the single dataset. However, the strategy of Davidov et al. (2026) sets static censoring times that are fixed before any interaction occurs and are not updated over time, even as the interaction sequence unfolds. Consequently, this static method is not ideal for our multi-turn setup. Furthermore, it often under-utilizes the budget; when an event is encountered before the budget is fully consumed, the trajectory terminates, and the remaining budget dedicated to that sample is lost: it cannot be redistributed to other samples.

However, dynamically allocating the censoring times is mathematically challenging. First, different sequences cannot actively share information or coordinate their decisions to stop or continue, since such communication breaks the independence assumptions required for this theoretical guarantee. Second, the online adaptation of C_i must respect the shared expected budget, even though future trajectories are unknown. Third, beyond validity, we should allocate this budget efficiently.

Importantly, we view our framework as a statistical tool for quantifying time-to-event under budget constraints. Predictive bounds can support safety assessment, model evaluation, and decision-making under budget constraints. For example, LLM developers could apply expensive audit functions, use a more conservative system prompt, or impose stricter sampling limits for prompts having a low LPB for the time-to-unsafe. Our framework is illustrated in Figure 1.

1.1 OUR CONTRIBUTION

In this work, we introduce *Dynamic Allocation via PRojected Optimization* (**DAPRO**), a novel paradigm that treats budget allocation as a sequential decision to form a statistically valid bound for the time-to-event. Specifically, our main contributions are:

- **Tighter Theoretical Guarantees**: We prove that DAPRO attains the desired coverage level with a novel bound that scales with the *mean* inverse-censoring weight only on a subset of trajectories rather than the *maximum* weight of all trajectories, leading to tighter guarantees than prior methods (Davidov et al., 2026; Gui et al., 2024). Crucially, our proof technique also removes the restrictive assumption that the censoring time C and true event time T must be conditionally independent given the prompt X. This result is of independent interest as it applies to a broad class of survival analysis tasks.
- **Budget Validity**: We formally prove that DAPRO satisfies the expected budget constraint.
- **Population-Level Performance Evaluation Under Censoring**: DAPRO can efficiently evaluate AI systems over a fixed benchmark: it provides unbiased estimates of metrics over the benchmark, such as the unsafe event rate, substantially reducing the costs of LLM API calls. Notably, this task does not require the i.i.d. assumption, since the metrics are evaluated directly over the benchmark of interest. Our experiments show that DAPRO reduces estimator variability and lowers the risk that a model appears safer than it is.
- **Comprehensive Empirical Evaluation**: We validate DAPRO across diverse tasks, including agentic task success, toxic content, LLM jailbreaks, and RAG-based hallucinations, demonstrating that our proposal generally achieves empirical coverage closer to the nominal level and utilizes the budget more efficiently compared to existing baselines.

1.2 PROBLEM SETUP

We consider a collection of base objectives $\{V_i\}_{i=1}^N$. In safety setups, an objective may correspond to a harmful intent, e.g., $V_i =$ "How can I bug someone's phone?", whereas in an agentic utility setting it may represent a task such as "Find a bug in the code." For each objective, an attacker or agent model $\mathcal{A}$ generates an initial prompt $X_i = \mathcal{A}(V_i)$. For instance, an attacker might wrap the objective in a roleplay persona, while an agent may implement a unit test. We define the resulting calibration dataset as $\mathcal{D}_{\text{cal}} = \{X_i\}_{i=1}^N$. We assume that the training data $\mathcal{D}_{\text{train}}$, used to fit the predictive model estimating $T \mid X$, has a separate budget. Importantly, our framework does not require fully observed training conversations: any predictor trained independently of the calibration set may be used, including one trained on censored data, data from another distribution, or external model predictions (Cunningham et al., 2026). Furthermore, the predictor can be pretrained once and reused across re-calibrations over different model configurations or benchmarks.

For each prompt X_i in $\mathcal{D}_{\text{cal}}$, the model $\mathcal{A}$ interacts sequentially with a target model $\mathcal{G}$ to elicit an event of interest. At step $t = 1$, the history is initialized with the initial prompt: $H_i(1) = \{A_i(1)\}$, where $A_i(1) := X_i$. At any step $t \geq 1$, the target model processes this prompt and outputs a response: $R_i(t) = \mathcal{G}(A_i(t))$. Next, we employ an audit function $\mathcal{J}$ to obtain a binary label $Y_i(t) = \mathcal{J}(A_i(t), R_i(t)) \in \{0, 1\}$, where $Y_i(t) = 1$ indicates that the event of interest has occurred (e.g., bug found), and $Y_i(t) = 0$ otherwise (e.g., task failed). If the key event occurred ($Y_i(t) = 1$), we terminate this process. Otherwise, we continue to step $t + 1$, where $\mathcal{A}$ generates a refined prompt $A_i(t + 1) = \mathcal{A}(H_i(t) \cup \{(R_i(t), Y_i(t))\})$ based on the target's previous responses. We update the interaction history: $H_i(t + 1) = H_i(t) \cup \{(R_i(t), Y_i(t), A_i(t + 1))\} \in \mathcal{H}_{t+1}$.

We define the uncensored time-to-event as the minimum number of steps required to achieve the event: $T_i = \min\{t \geq 1 : Y_i(t) = 1\}$. The censored time-to-event is thus $\tilde{T}_i = \min(T_i, C_i)$. For notational convenience, we denote the complete interaction history upon termination as $H_i = \{H_i(t)\}_{t=1}^{\min\{t_{\max}, T_i\}}$, where $t_{\max}$ is the maximal sequence length. If the allocated budget exceeds the true event time ($C_i > T_i$), the interaction naturally terminates at T_i, and no further budget is consumed. Therefore, the expected calibration budget constraint can be formulated as:

$$\mathbb{E}\left[\sum_{i=1}^{N} \tilde{T}_i\right] \leq B. \tag{1}$$

Our goal is to construct an LPB $\hat{L}(X_{\text{test}})$ given an initial test prompt X_{test} for the test event time T_{test}. Importantly, at test time, we construct the LPB *without* running any LLM exchanges, relying solely on the initial prompt X_{test}. This design choice reflects two practical considerations. First, a test-time budget may be extremely limited or entirely unavailable. Second, when the target model operates autonomously and can act on external systems, assessing a prompt before the interaction begins can be valuable, since unsafe actions taken during the interaction may be difficult to reverse. Finally, given a target coverage level $1-\alpha \in (0,1)$ and a tolerance level $\delta \in (0,1)$, the LPB should satisfy:

$$\mathbb{P}\left(\hat{L}(X_{\text{test}}) \leq T_{\text{test}} \mid \mathcal{D}_{\text{all}}\right) \geq 1-\alpha, \tag{2}$$

with probability at least $1-\delta$ over the randomness of the data $\mathcal{D}_{\text{all}} = \mathcal{D}_{\text{train}} \cup \mathcal{D}_{\text{cal}}$. We remark that the probability in equation 2 is taken over random draws of $X_{\text{test}}, T_{\text{test}}$. An LPB $\hat{L}$ satisfying this requirement is called a Probably Approximately Correct (PAC) LPB at level α with tolerance δ.

2 BACKGROUND

Our approach builds on the framework of Davidov et al. (2026), which refines conformalized survival analysis (Gui et al., 2024) to construct a valid LPB for the time-to-unsafe under a known censoring mechanism. We outline the components relevant for our dynamic censoring mechanism.

Suppose we are given quantile estimates $\hat{q}_\tau(x)$ for the time-to-event from a pre-trained model. Since the raw estimates $\hat{q}_\alpha(x)$ might not be sufficiently accurate, using them directly as LPBs might not achieve $1-\alpha$ coverage. Thus, we tune a quantile level $\hat{\tau}$ such that the resulting LPB $\hat{q}_{\hat{\tau}}(x)$ satisfies the coverage requirement. We obtain this calibrated quantile level by weighting the errors of the raw quantile estimates to account for the distribution shift induced by the censoring:

$$\hat{\alpha}(\tau) = \frac{1}{|\mathcal{I}_{\text{cal}}|} \sum_{i \in \mathcal{I}_{\text{cal}}} w_\tau(i)\, \mathbb{I}\{\tilde{T}_i < \hat{q}_\tau(X_i) \leq C_i\}, \tag{3}$$

where the weights are formulated as: $w_\tau(i) = \mathbb{P}\big[\hat{q}_\tau(X_i) \leq C_i \;\big|\; \{X_j\}_{j \in \mathcal{I}_{\text{cal}}}\big]^{-1}$. Notice that these weights are known since the censoring mechanism is actively designed by the algorithm. The calibrated LPB is then $\hat{L}(x) = \hat{q}_{\hat{\tau}}(x)$, where $\hat{\tau}$ is defined as the largest level achieving valid coverage:

$$\hat{\tau} = \sup\big\{\tau \in \mathcal{T} : \sup_{\tau' \leq \tau} \hat{\alpha}(\tau') \leq \alpha\big\}. \tag{4}$$

Above, $\mathcal{T}$ is the search space for τ. As established by Davidov et al. (2026); Gui et al. (2024), this LPB holds a PAC-type coverage guarantee that, informally, includes two terms: (i) the desired coverage level $1-\alpha$, as in equation 2; and (ii) a slack term that scales with the worst-case maximal weight $w_\tau(i)$. Further, Davidov et al. (2026) show that the variance of the LPBs is governed by the *average* weight; however, this connection was only justified under restrictive oracle assumptions.

To control the value of the maximal weight and bound the number of LLM interactions, Davidov et al. (2026) introduce two refinements that we adopt.

First, the raw LPB estimates are capped at $M \leq t_{\max}$, yielding trimmed estimates $\hat{f}_\tau(x) := \min(\hat{q}_\tau(x), M)$. By replacing $\hat{q}$ with $\hat{f}$ in equation 3, this trimming bounds the maximal weight, affecting the tightness of the PAC-type coverage guarantee. From an evaluation perspective, the trimming allows us to assess the coverage rate of the LPB. With this trimming, we can always observe whether the interaction survives past the predicted LPB.

The second refinement that we adopt from Davidov et al. (2026) is the use of a fixed quantile level τ for the budget allocation. In more detail, recall from equation 3 that a conversation i contributes to the miscoverage estimator $\hat{\alpha}(\tau)$ only when $\hat{f}_\tau(X_i) \leq C_i$. To allocate the budget efficiently, we aim to maximize the number of samples satisfying this condition. However, since the calibrated quantile $\hat{\tau}$ equation 4 is unknown prior to budget allocation, we follow Davidov et al. (2026) and fix a prior quantile level τ_{prior}, reflecting our belief that $\hat{\tau} \in [0, \tau_{\text{prior}}]$. For example, with $1-\alpha = 90\%$ and a reasonably accurate model, $\tau_{\text{prior}} = 30\%$ is an appropriate choice, as the α-quantile is expected to lie well below 30%. Despite using a fixed τ_{prior}, Davidov et al. (2026) showed that it is still possible to construct valid weights $w_\tau(i) = \mathbb{P}(\hat{f}_\tau(X_i) \leq C_i)^{-1}$ for all $\tau \leq \tau_{\text{prior}}$, which is required to compute equation 3. This procedure is summarized in Algorithm 1 in Appendix D.1.

Due to space constraints, we include an extended discussion of additional related works in conformal prediction, survival analysis, and LLM evaluation in Appendix A.

3 Proposed Method

Our adaptive budget allocation strategy operates in two phases. Before these phases, we randomly partition the calibration points, indexed by $\mathcal{I}_{\text{cal}} = \{1, \ldots, N\}$, into three disjoint sets: $\mathcal{I}_{\text{cal1}}$, $\mathcal{I}_{\text{crc}}$, and $\mathcal{I}_{\text{cal2}}$. We re-index the samples in the first split as $i \in \{1, \ldots, N_1\}$ for notational convenience.

In the first phase, we learn an optimized acquisition policy on $\mathcal{I}_{\text{cal1}}$. Our allocation strategy relies on a scoring function $\mathcal{S}_t : \mathcal{H}_t \to \mathbb{R}$ that maps the conversation trajectory observed up to time t to a real-valued score. A higher score indicates a stronger signal to spend a budget unit and advance to the next step. For example, this score could be the estimated probability of observing an event at the current step. This scoring function can be trained on the training data $\mathcal{D}_{\text{train}}$. Our proposed policy operates iteratively: at each step t, we compute the current score $S_i(t)$ and map it to a continuation probability $P_i(t) \in [0, 1]$. Then, with probability $P_i(t)$, we advance to the next step; otherwise, the acquisition for this sample is terminated. We aim to optimize this score-to-probability mapping over $\mathcal{I}_{\text{cal1}}$ to reduce the variance of the weighted miscoverage estimator equation 3 while satisfying an expected budget constraint. This allows us to allocate more budget to trajectories that are more likely to reduce the variance of the miscoverage, while terminating less informative conversations early. After learning the mapping, we leverage $\mathcal{I}_{\text{crc}}$ to calibrate its expected budget with conformal risk control (CRC) (Angelopoulos et al., 2024). In the second phase, we deploy this policy on the remaining indices $\mathcal{I}_{\text{cal2}}$ to acquire new interactions. By leveraging the interactions observed in $\mathcal{I}_{\text{cal1}}$, we selectively terminate conversations that are unlikely to contribute to reducing the LPB variance.

3.1 Phase I: Learning the Optimized Acquisition Policy on $\mathcal{I}_{\text{cal1}}$

To learn an optimized mapping from scores to probabilities, we must first expend budget to observe the full conversation history for all samples in $\mathcal{I}_{\text{cal1}} \cup \mathcal{I}_{\text{crc}}$, until we observe an event, or reach $\hat{f}_{\text{prior}}(X_i)$, where $\hat{f}_{\text{prior}}(\cdot) := \hat{f}_{\tau_{\text{prior}}}(\cdot)$. For all $i \in \mathcal{I}_{\text{cal1}}$, we set the censoring times as $C_i = \hat{f}_{\text{prior}}(X_i)$, and expend $b_i := \min(T_i, \hat{f}_{\text{prior}}(X_i))$ budget units. We require that the total budget B is sufficient for this observation phase, such that $B > \sum_{i \in \mathcal{I}_{\text{cal1}} \cup \mathcal{I}_{\text{crc}}} b_i$. The budget remaining for Phase II and the CRC is therefore $B_{\text{rem}} := \left(B - \sum_{i \in \mathcal{I}_{\text{cal1}}} b_i\right)$ and the per-sample budget is $\bar{B}_2 := \frac{1}{|\mathcal{I}_{\text{crc}}| + |\mathcal{I}_{\text{cal2}}|} B_{\text{rem}}$.

After annotating the data in $\mathcal{I}_{\text{cal1}}$, we compute the scores for all observed time steps: $S_i(t) = \mathcal{S}_t(H_i(t))$. The inverse-probability weights are given by $w(i) = \mathbb{P}(\hat{f}_{\text{prior}}(X_i) \leq C_i)^{-1} = \left(\prod_{t=1}^{b_i} P_i(t)\right)^{-1}$. As shown in Section 3.3 and Appendix C.1, for a fixed candidate τ, the variance of the weighted miscoverage estimator in equation 3 scales with the mean of weighted errors $\mathbb{E}\Big[w(i)\mathbb{I}\{T_i < \hat{f}_\tau(X_i)\}\Big] = \mathbb{E}\Big[w(i)\sum_{t < \hat{f}_\tau(X_i)} \mathbb{I}\{T_i = t\}\Big]$. Ideally, we would minimize this quantity at the calibrated level $\hat{\tau}$ from equation 4. However, $\hat{\tau}$ is selected only after data acquisition and is therefore unavailable when learning the acquisition policy. We instead use $\tau = \alpha$ as a fixed candidate level, since it is the target miscoverage level. This focuses the allocation on trajectories that can contribute at the target level, rather than the more conservative choice $\tau = \tau_{\text{prior}}$, which considers errors over the entire candidate range $\mathcal{T}$. Thus, we aim to reduce the weights of trajectories satisfying $T_i < \hat{f}_\alpha(X_i)$. Directly optimizing with this indicator $\mathbb{I}\{T_i < \hat{f}_\alpha(X_i)\}$, however, can be unstable when $\mathcal{I}_{\text{cal1}}$ is small. We therefore use their estimated probabilities as a smooth proxy, which provides a stronger signal for the optimization. Our goal is to learn a monotone score-to-probability mapping M_t^{raw} for each turn t:

$$\begin{aligned} \min_{M^{\text{raw}}} \quad & \frac{1}{N_1} \sum_{i=1}^{N_1} \sum_{t < \hat{f}_\alpha(X_i)} \frac{\widehat{\mathbb{P}}(T_i = t \mid T_i \geq t, H_i(t))}{\prod_{s=1}^{t} M_s^{\text{raw}}(S_i(s))} \\ \text{s.t.} \quad & \frac{1}{N_1} \sum_{i=1}^{N_1} \mathcal{B}(M^{\text{raw}}(S_i)) \leq \bar{B}_2, \quad S_i(t) \leq S_j(t) \Longrightarrow M_t^{\text{raw}}(S_i(t)) \leq M_t^{\text{raw}}(S_j(t)). \end{aligned} \tag{5}$$

Here, $\mathcal{B}(P_i)$ is the expected budget expended on the i-th sample. Since each interaction step consumes one budget unit, and the probability of the conversation surviving to step t is the product of all prior continuation probabilities $\prod_{j=1}^{t} P_i(j)$, the expected budget is the sum of these survival probabilities: $\mathcal{B}(P_i) = \sum_{t=1}^{b_i} \prod_{j=1}^{t} P_i(j)$. We fully describe how we solve this optimization in Appendix D.3. Thus, we optimize a tractable and stable proxy motivated by the exact variance expression. The ablation in Appendix F.4.4 shows that this optimization directly improves the variance.

If $\{M_t^{\text{raw}}\}_{t=1}^{t_{\max}}$ generalize well to unseen trajectories, then the resulting policy satisfies the expected budget constraint, as formally shown in Appendix C.3. Empirically, we find this typically holds: even without additional calibration, DAPRO satisfies the target budget across most experiments. For settings that require a stronger guarantee, we introduce an optional second calibration step based on conformal risk control (CRC) (Angelopoulos et al., 2024), which guarantees expected budget control without relying on accurate generalization of the raw policy. Specifically, we consider a one-parameter family of mappings $M_t(s;\lambda)$ such that the induced expected budget is nondecreasing in λ. A simple choice is $M_t(s;\lambda) := \lambda \cdot M_t^{\text{raw}}(s)$. To calibrate λ, we fully observe each trajectory $i \in \mathcal{I}_{\text{crc}}$ until either the event occurs or $\hat{f}_{\text{prior}}(X_i)$ is reached, consuming $b_i = \min\{T_i, \hat{f}_{\text{prior}}(X_i)\}$ budget units. These trajectories allow us to evaluate the expected deployment cost $\mathcal{B}(M(S_i;\lambda))$ that the candidate policy would incur. Before CRC calibration, we cap the candidate family so that its conditional expected cost on every trajectory is at most $C_{\max}$. This prevents CRC from resulting in a calibrated policy that is too conservative. We then select the largest value of λ whose CRC upper bound on the expected per-sample cost remains within the available budget:

$$\widehat{\lambda} := \sup\left\{ \lambda \in [0,1] : \frac{\sum_{i\in\mathcal{I}_{\text{crc}}} \left(\mathcal{B}(M(S_i;\lambda)) + \frac{|\mathcal{I}_{\text{crc}}|}{|\mathcal{I}_{\text{cal2}}|} b_i\right) + C_{\max} + \frac{|\mathcal{I}_{\text{crc}}|}{|\mathcal{I}_{\text{cal2}}|} t_{\max}}{|\mathcal{I}_{\text{crc}}| + 1} \le \frac{B_{\text{rem}}}{|\mathcal{I}_{\text{cal2}}|} \right\}. \tag{6}$$

Finally, the calibrated continuation policy is formulated as $M_t(\cdot) := M_t(\cdot;\hat{\lambda})$. Thus, the learned mappings M_t^{raw} determine the relative allocation of budget across trajectories, while CRC calibrates the overall budget consumption to obtain an expected budget guarantee.

3.2 Phase II: Adaptive Acquisition on $\mathcal{I}_{\text{cal2}}$

Armed with the sequence of score-to-probability mapping models $\{M_t\}_{t=1}^{t_{\max}}$, we adaptively acquire interactions for the second split, $\mathcal{I}_{\text{cal2}}$. We initialize the process at $t = 1$ and define a continuation indicator $\xi_i(0) = 1$. At each step t, we compute the score $S_i(t) = \mathcal{S}_t(H_i(t))$ and transport it into a continuation probability $P_i(t) = M_t(S_i(t))$. We then draw a Bernoulli random variable with parameter $P_i(t)$ to determine if the interaction should proceed. Formally, at each step t, we set $P_i(t) = M_t(S_i(t))$, draw $Z_t^{(i)} \sim \text{Bern}(P_i(t))$, and update $\xi_i(t) = \xi_i(t-1)Z_t^{(i)}$. If $\xi_i(t) = 1$, we advance to the next step: we acquire the new conversation exchange $H_i(t+1)$, set $t \leftarrow t+1$, and repeat the process. The interaction terminates when $\xi_i(t) = 0$ or when reaching the predefined boundary $t = \min(\hat{f}_{\text{prior}}(X_i), T_i)$.

Finally, we define the censoring time for each sample $i \in \mathcal{I}_{\text{cal2}}$ as: $C_i = \hat{f}_{\text{prior}}(X_i)$ if $\xi_i(\min\{T_i, \hat{f}_{\text{prior}}(X_i)\}) = 1$; otherwise, C_i is the last successfully reached step. Consequently, for samples $i \in \mathcal{I}_{\text{cal2}}$ that successfully reach their target sequence length, i.e., $\hat{f}_{\text{prior}}(X_i) \le C_i$, the probability of achieving this event is the product of the sequential continuation probabilities:

$$w(i)^{-1} = \mathbb{P}(\hat{f}_{\text{prior}}(X_i) \le C_i \mid X_i, H_i, T_i, \{(X_j, H_j, T_j)\}_{j\in\mathcal{I}_{\text{cal1}}\cup\mathcal{I}_{\text{crc}}}, \mathcal{D}_{\text{train}}) = \prod_{t=1}^{\min(\hat{f}_{\text{prior}}(X_i), T_i)} P_i(t).$$

For samples $i \in \mathcal{I}_{\text{cal1}} \cup \mathcal{I}_{\text{crc}}$ we set $w(i) = 1$. Lastly, we employ Algorithm 1 using the samples we acquired $\{(C_i, w(i), \tilde{T}_i)\}_{i\in\mathcal{I}_{\text{cal}}}$. For convenience, we summarize this procedure in Algorithm 3 in Appendix D.3. In what follows, we prove that this process satisfies the coverage requirement and expected budget constraint.

3.3 THEORETICAL GUARANTEES

First, we show that the LPB constructed by DAPRO holds a PAC-type coverage guarantee.[1]

Theorem 1 (Coverage validity (informal)). *Fix a miscoverage level $\alpha \in (0,1)$ and a tolerance level $\delta \in (0,1)$. Suppose that $\{(X_i, T_i, H_i)\}_{i\in\mathcal{I}_{\text{cal}}}$ and $(X_{test}, T_{test}, H_{\text{test}})$ are drawn i.i.d., and that $\hat{f}_\tau(x)$ is non-decreasing. Further assume $\mathbb{E}[w(i)\mathbb{I}\{T_i < \hat{f}_{\text{prior}}(X_i)\} \mid \{(X_i, H_i, T_i)\}_{i\in\mathcal{I}_{\text{cal1}}\cup\mathcal{I}_{\text{crc}}}, \mathcal{D}_{\text{train}}] \le \bar{\nu}, \forall i \in \mathcal{I}_{\text{cal2}}$ for $\bar{\nu} > \alpha^2$ a.s. Then, with probability at least $1-\delta$ over the draw of $\mathcal{D}_{\text{all}}$, the LPB $\hat{L}(x)$ generated by DAPRO satisfies:*

$$\mathbb{P}\Big[T_{test} \ge \hat{L}(X_{test}) \,\Big|\, \mathcal{D}_{\text{all}}\Big] \;\ge\; 1-\alpha-\frac{\log(1/\delta)}{3|\mathcal{I}_{\text{cal}}|}-\sqrt{\frac{\log^2(1/\delta)}{9|\mathcal{I}_{\text{cal}}|^2}+\frac{2\big(\bar{\nu}-\alpha^2\big)\log(1/\delta)}{|\mathcal{I}_{\text{cal}}|}}.$$

Importantly, this coverage guarantee holds in finite samples for any quantile estimator $\hat{q}_\tau(\cdot)$, any LLM, and any data distribution. We defer the proof of Theorem 1 to Appendix C.1. We remark that Phase I uses $\hat{f}_\alpha$ only as a fixed reference predictor for learning the allocation policy. The calibrated predictor $\hat{f}_{\hat{\tau}}$ selected afterwards may differ. The theorem therefore considers the more conservative event $T_i < \hat{f}_{\text{prior}}(X_i)$: every candidate predictor $\hat{f}_\tau(X_i)$ with $\tau \le \tau_{\text{prior}}$ satisfies $\mathbb{I}\{T_i < \hat{f}_\tau(X_i)\} \le \mathbb{I}\{T_i < \hat{f}_{\text{prior}}(X_i)\}$. Hence, $\bar{\nu}$ bounds the weighted errors for any predictor the calibration may select.

While this proof builds on the theoretical foundations in (Gui et al., 2024; Davidov et al., 2026), we highlight three key implications that distinguish our result:

- **Relaxed Assumptions:** Crucially, unlike prior conformal frameworks (Davidov et al., 2026; Gui et al., 2024), our result does not rely on the assumption of conditional independence between the censoring times C and event times T given the covariates X. DAPRO inherently violates this assumption by using T for early stopping, making this relaxation necessary.
- **Tighter Scaling:** Our coverage gap scales as $\mathcal{O}(\sqrt{\bar{\nu}/N})$ where $\bar{\nu}$ is the mean of inverse-probability weights only for trajectories that contribute to the miscoverage estimator: $w(i)\mathbb{I}\{T_i < \hat{f}_{\text{prior}}(X_i)\}$. In contrast, prior guarantees (Gui et al., 2024; Davidov et al., 2026) depend on the worst-case weight $w_{\max} \ge w(i)\mathbb{I}\{T_i < \hat{f}_{\text{prior}}(X_i)\}$. Thus, our coverage gap can remain small even when some inverse-probability weights are large. We illustrate the efficiency of our coverage gap in Figure 20 in Appendix F.3.
- **Independent Interest:** Our proof technique can be directly applied to the static procedure of Davidov et al. (2026) and to a broad class of survival analysis tasks.

We show that DAPRO satisfies the budget constraint in equation 1; see Appendix C.2 for the proof.

Theorem 2 (Budget validity). *The expected budget used by DAPRO is upper-bounded by B.*

4 EXPERIMENTS

In this section, we evaluate DAPRO on three sequential safety and utility tasks. For safety, we generate attacker-target interactions using PAIR (Chao et al., 2025) from 10,000 prompts sampled from RealToxicityPrompts (Gehman et al., 2020), where the event is a toxic response, and from the Anthropic Red-Team dataset (Bai et al., 2022; Ganguli et al., 2022), where the event is a successful jailbreak. To evaluate agentic utility, we use 10,000 examples from AutoIF (Dong et al., 2025), a benchmark for complex instruction following. Here, the "attacker" is replaced by a helper agent, and the event occurs when the target model successfully satisfies all instructions.

To generate the interactions, we utilize Qwen 2.5 14B Instruct (Team, 2024) as the attacker model, as its outputs had the lowest formatting error rate among the models we examined. For the target models, we use Qwen 2.5 14B Instruct, Llama 3.1 8B Instruct (Grattafiori et al., 2024), Phi 4 Mini Instruct (Abouelenin et al., 2025), and Gemma 3 4B Instruct (Team, 2025). We define the target event ($Y(t) = 1$) using an audit function: for the Toxicity dataset, following (Davidov et al., 2026),

[1] An analogous UPB guarantee holds; see Appendix B.

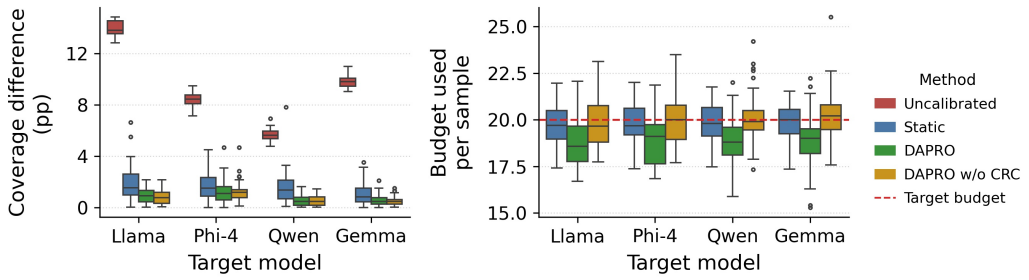

Figure 2: **Toxicity** dataset: absolute coverage deviation in percentage points (pp) and average budget utilized across four target LLMs. Target coverage rate: 90% and target $\bar{B} = 20$ budget per sample.

we use Detoxify (Hanu & Unitary team, 2020) with a threshold of 0.5 to identify toxic responses. For the Red-Team dataset, we employ Llama Guard as the judge (Inan et al., 2023). Finally, for the AutoIF dataset, we verify a successful instruction-following response programmatically. Additionally, Appendix F.1.4 evaluates DAPRO on a RAG-based dataset, in which the event is a hallucination. We cap the maximum sequence length at $t_{\max} = 200$. The full details regarding the dataset generation are provided in Appendix E.1.

We split the data into training (4000), calibration (3000), and test (3000) sets. We fix the training data and fit a transformer as the predictive model, estimating $\hat{q}_\tau$ over it, using the full conversation history. In our experiments, our training data is fully observed up to time $t_{\max}$, although our framework allows any predictor trained independently of the calibration set to be used. In Appendix F.4.10 we show that our framework can still provide informative estimates even with smaller training budgets. At any given time t_1, the model estimates the probability of an event occurring at any future time step $t_2 > t_1$. All methods use the same predictive models. In all experiments, we evaluate the methods over 50 independent random splits of the calibration and test sets. During each trial, the calibration algorithms are employed using the calibration data and evaluated on the test set. Following (Davidov et al., 2026), we set $\tau_{\text{prior}} = 0.56$, and $M = 200$. We compare our proposal to an uncalibrated baseline, which constructs the LPB directly from the raw model predictions at the target α level, $\hat{q}_\alpha$, and to the static optimized allocation strategy of Davidov et al. (2026).

Figure 2 presents the absolute deviation from the target 90% coverage rate and average budget per sample used by each method on the Toxicity dataset. This figure reveals that DAPRO satisfies the expected budget constraint while attaining the lowest coverage deviation. Specifically, averaged across the four target models, DAPRO achieved $1.90\times$ lower mean absolute deviation from the target coverage and $3.32\times$ lower coverage-rate variance than the static baseline. In addition, Figure 3 details the same metrics on the AutoIF dataset for both an LPB (target coverage 90%, $\bar{B} = B/|\mathcal{I}_{\text{cal}}| = 10$) and a UPB (target coverage 80%, $\bar{B} = 20$). We evaluate the UPB at this lower target coverage since a 90% target pushes the bound too close to the maximum $t_{\max} = 200$, which artificially reduces the variance differences between methods. Across both bound types, the results reveal the same trend: DAPRO attains a coverage rate closer to the nominal level compared to the baselines.

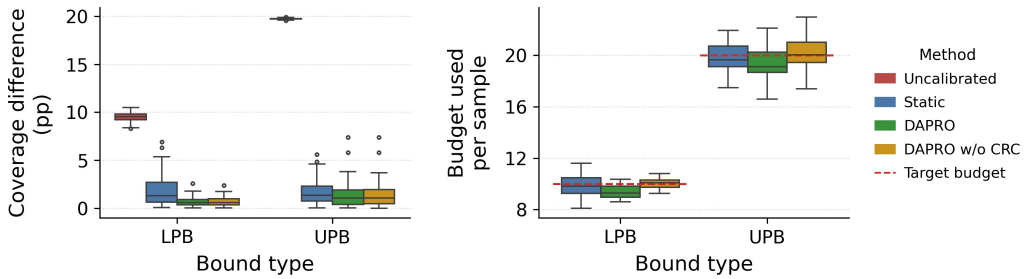

Figure 3: **AutoIF** dataset: coverage deviation (left) and budget utilized (right) over the Qwen 2.5 target model. Target coverage rate: 90% for LPB and 80% for UPB.

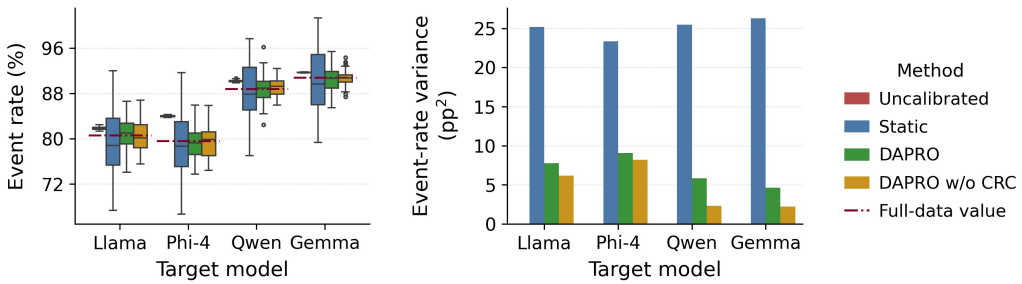

Figure 4: **Red-Team event-rate estimation**. Left: estimated population event rates. Right: variance of the event-rate estimator. Target $\bar{B} = 20$ budget per sample.

We evaluate the jailbreak rate on the Red-Team benchmark. Figure 4 shows that while both DAPRO and the static baseline closely match the full-data event rate on average, DAPRO exhibits approximately 72% lower variance on average. This improvement is also reflected in the worst observed estimates: for Phi-4 Mini, whose full-data event rate is 79.6%, the lowest estimate under Static is 66.6%, compared to 73.7% under DAPRO. Thus, DAPRO more reliably estimates the full-data event rate. The lower variance of DAPRO reduces the risk that a low estimate will make an LLM appear safer than it is. More details on this metric-estimation experiment are given in Appendix F.2.

Additional experimental setups and diagnostics are deferred to Appendix F. Furthermore, our ablation study in Appendix F.4 reveals that DAPRO is robust to degradation of its components and maintains a tight coverage rate while satisfying the expected budget constraint.

5 Discussion and Impact Statement

We introduced DAPRO, a novel dynamic budget allocation method for LLM evaluation. Our approach can be employed to construct reliable LPBs for individual prompts and extract unbiased metrics over benchmarks, such as the jailbreak rate. By updating censoring times online based on ongoing attacker-target interactions, our method achieves lower variance, as supported by our PAC-type coverage guarantee, which is tighter than prior approaches. Our comprehensive experiments on real datasets validate these theoretical claims, demonstrating tighter bounds compared to baselines.

While our method has an expected budget guarantee, our experiments show that it often underutilizes the budget due to the CRC calibration. So, for applications that do not require a finite-sample budget guarantee, practitioners can omit the CRC step—our experiments show it generally satisfies the budget constraint despite the theoretical requirement for an accurate model. Furthermore, the calibrated quantile level $\hat{\tau}$ in our current method is a single scalar applied identically to all test prompts. An intriguing extension would be to make $\hat{\tau}$ adaptive to the test prompt X_{test}, possibly leveraging the newly acquired calibration data to select a prompt-specific quantile level, e.g., by utilizing ideas from (Feldman et al., 2021; Gibbs et al., 2025). This would allow the LPB to be tighter for predictable interactions and more conservative for those with high variance. Another direction may be engineering a score that better represents the safety or utility of a prompt, which could be mapped to a better continuation probability that leads to a lower variance.

We note that our coverage guarantees rely on the assumption that the calibration and test sets are i.i.d. In practice, the data distribution may shift over time as models are updated or new attack strategies emerge. Developing methods that are robust to such distributional shifts (Gibbs & Candes, 2021; Feldman et al., 2023) or audit annotation errors (Einbinder et al., 2024; Penso & Goldberger, 2024; Feldman & Romano, 2024; Feldman et al., 2026) is an important direction for future work. Importantly, this i.i.d. assumption is not required for metric estimation over a fixed benchmark, since these metrics are evaluated directly on the benchmark of interest rather than generalized to unseen prompts. Finally, while our framework provides statistical guarantees to help developers evaluate AI systems and set usage limits, its insights into model failures could potentially be exploited by adversaries, akin to many developments in ML. Therefore, our proposal must be deployed as a complementary tool within a broader safety, alignment, and governance pipeline.

AI USE STATEMENT

In this work, we used generative AI to generate data sets, assist in reviewing and polishing our writing and theoretical proofs, provide feedback on research methodology and experiments, assist in the writing of proofs, implement methods in code, and improve the readability of our paper. We have not used generative AI tools to help develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, propose or refine hypotheses, assist with translation, clean and reformat datasets, support qualitative and thematic data analysis, interpret results, create or modify scientific figures or images, suggest experimental parameters, format references, or suggest structure for the paper. Formulating questions for surveys or interviews, and the creation of artifacts are not applicable to this work. Additionally, we used generative AI tools to create or edit software code, including implementing methods in code, identifying relevant literature, and proposing a title or keywords for our paper. LLM-generated code was verified and tested for correctness by human review, experimental results, and automatic tests. The authors checked all theoretical results, proofs, experimental conclusions, and references as well. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

The experimental setup, including model architectures, training procedures, hyperparameters, evaluation protocols, and data generation, is described in Appendix E. Additional implementation details are provided in Appendix D.3 and further ablations are reported in Appendix F.4. Code implementing our method and the data-generating process necessary to reproduce the experiments are included in the supplementary material.

ETHICS STATEMENT

The techniques developed in this work are intended to improve evaluation of sequential AI systems. However, the proposed method could potentially be adapted for exploiting failures of LLMs. We have therefore discussed the potential negative implications and the relevant limitations of our method in Section 5. We believe that the scientific benefits of introducing our method and results outweigh the identified risks, while acknowledging that additional safeguards may be necessary in deployment settings.

REFERENCES

Abdelrahman Abouelenin, Atabak Ashfaq, Adam Atkinson, Hany Awadalla, Nguyen Bach, Jianmin Bao, Alon Benhaim, Martin Cai, Vishrav Chaudhary, Congcong Chen, et al. Phi-4-mini technical report: Compact yet powerful multimodal language models via mixture-of-loras. *arXiv preprint arXiv:2503.01743*, 2025.

Eitan Altman. *Constrained Markov decision processes*. Routledge, 2021.

Anastasios N Angelopoulos, Stephen Bates, Clara Fannjiang, Michael I Jordan, and Tijana Zrnic. Prediction-powered inference. *Science*, 382(6671):669–674, 2023.

Anastasios Nikolas Angelopoulos, Stephen Bates, Adam Fisch, Lihua Lei, and Tal Schuster. Conformal risk control. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=33XGfHLtZg.

Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, et al. Qwen technical report. *arXiv preprint arXiv:2309.16609*, 2023.

Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, et al. Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*, 2022.

Stephen Bates, Anastasios Angelopoulos, Lihua Lei, Jitendra Malik, and Michael Jordan. Distribution-Free, Risk-Controlling Prediction Sets. *Journal of the ACM (JACM)*, 68(6):1–34, 2021. doi: 10.1145/3478535.

Stephen Bates, Emmanuel Candès, Lihua Lei, Yaniv Romano, and Matteo Sesia. Testing for outliers with conformal p-values. *The Annals of Statistics*, 51(1):149–178, 2023.

Emmanuel Candès, Lihua Lei, and Zhimei Ren. Conformalized survival analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(1):24–45, 2023. doi: 10.1093/jrsssb/qkac004.

Patrick Chao, Alexander Robey, Edgar Dobriban, Hamed Hassani, George J Pappas, and Eric Wong. Jailbreaking black box large language models in twenty queries. In *2025 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*, pp. 23–42. IEEE, 2025.

Hoagy Cunningham, Jerry Wei, Zihan Wang, Andrew Persic, Alwin Peng, Jordan Abderrachid, Raj Agarwal, Bobby Chen, Austin Cohen, Andy Dau, et al. Constitutional classifiers++: Efficient production-grade defenses against universal jailbreaks. *arXiv preprint arXiv:2601.04603*, 2026.

Sanjoy Dasgupta, Daniel J Hsu, and Claire Monteleoni. A general agnostic active learning algorithm. *Advances in neural information processing systems*, 20, 2007.

Hen Davidov, Shai Feldman, Gil Shamai, Ron Kimmel, and Yaniv Romano. Conformalized Survival Analysis for General Right-Censored Data. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=JQtuCumAFD.

Hen Davidov, Shai Feldman, Gilad Freidkin, and Yaniv Romano. Calibrated predictive lower bounds on time-to-unsafe-sampling in LLMs. In *The 29th International Conference on Artificial Intelligence and Statistics*, 2026.

Guanting Dong, Keming Lu, Chengpeng Li, Tingyu Xia, Bowen Yu, Chang Zhou, and Jingren Zhou. Self-play with execution feedback: Improving instruction-following capabilities of large language models. In *The Thirteenth International Conference on Learning Representations*, 2025.

Bat-Sheva Einbinder, Shai Feldman, Stephen Bates, Anastasios N Angelopoulos, Asaf Gendler, and Yaniv Romano. Label noise robustness of conformal prediction. *Journal of Machine Learning Research*, 25(328):1–66, 2024.

Valerii Vadimovich Fedorov. *Theory of optimal experiments*. Elsevier, 2013.

Shai Feldman and Yaniv Romano. Robust conformal prediction using privileged information. *Advances in Neural Information Processing Systems*, 37:117813–117852, 2024.

Shai Feldman, Stephen Bates, and Yaniv Romano. Improving conditional coverage via orthogonal quantile regression. *Advances in neural information processing systems*, 34:2060–2071, 2021.

Shai Feldman, Liran Ringel, Stephen Bates, and Yaniv Romano. Achieving risk control in online learning settings. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856.

Shai Feldman, Stephen Bates, and Yaniv Romano. Conformal prediction with corrupted labels: Uncertain imputation and robust re-weighting. In *International Conference on Learning Representations*, volume 2026, pp. 30901–30942, 2026.

Deep Ganguli, Liane Lovitt, Jackson Kernion, Amanda Askell, Yuntao Bai, Saurav Kadavath, Ben Mann, Ethan Perez, Nicholas Schiefer, Kamal Ndousse, et al. Red teaming language models to reduce harms: Methods, scaling behaviors, and lessons learned. *arXiv preprint arXiv:2209.07858*, 2022.

Samuel Gehman, Suchin Gururangan, Maarten Sap, Yejin Choi, and Noah A Smith. RealToxicityPrompts: Evaluating Neural Toxic Degeneration in Language Models. In *Findings of the Association for Computational Linguistics: EMNLP*, pp. 3356–3369, 2020.

Isaac Gibbs and Emmanuel Candes. Adaptive conformal inference under distribution shift. *Advances in Neural Information Processing Systems*, 34:1660–1672, 2021.

Isaac Gibbs, John J Cherian, and Emmanuel J Candès. Conformal prediction with conditional guarantees. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 87(4): 1100–1126, 2025.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

Yu Gui, Rohan Hore, Zhimei Ren, and Rina Foygel Barber. Conformalized survival analysis with adaptive cut-offs. *Biometrika*, 111(2):459–477, 2024. doi: 10.1093/biomet/asad076.

Jinyong Hahn, Keisuke Hirano, and Dean Karlan. Adaptive experimental design using the propensity score. *Journal of Business & Economic Statistics*, 29(1):96–108, 2011.

Laura Hanu and Unitary team. Detoxify. Github. https://github.com/unitaryai/detoxify, 2020.

Daniel G. Horvitz and Donovan J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685, 1952.

Hakan Inan, Kartikeya Upasani, Jianfeng Chi, Rashi Rungta, Krithika Iyer, Yuning Mao, Michael Tontchev, Qing Hu, Brian Fuller, Davide Testuggine, et al. Llama guard: Llm-based input-output safeguard for human-ai conversations. *arXiv preprint arXiv:2312.06674*, 2023.

Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. Survey of hallucination in natural language generation. *ACM computing surveys*, 55(12):1–38, 2023.

Carlos E Jimenez, John Yang, Alexander Wettig, Shunyu Yao, Kexin Pei, Ofir Press, and Karthik Narasimhan. Swe-bench: Can language models resolve real-world github issues? In *International Conference on Learning Representations*, volume 2024, pp. 54107–54157, 2024.

Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations*, 2015.

Jannik Kossen, Sebastian Farquhar, Yarin Gal, and Tom Rainforth. Active testing: Sample-efficient model evaluation. In *International Conference on Machine Learning*, pp. 5753–5763. PMLR, 2021.

Aounon Kumar, Chirag Agarwal, Suraj Srinivas, Aaron Jiaxun Li, Soheil Feizi, and Himabindu Lakkaraju. Certifying llm safety against adversarial prompting. *arXiv preprint arXiv:2309.02705*, 2023.

Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph Gonzalez, Hao Zhang, and Ion Stoica. Efficient memory management for large language model serving with pagedattention. In *Proceedings of the 29th symposium on operating systems principles*, pp. 611–626, 2023.

Yang Li, Jie Ma, Miguel Ballesteros, Yassine Benajiba, and Graham Horwood. Active evaluation acquisition for efficient llm benchmarking. *arXiv preprint arXiv:2410.05952*, 2024.

Anay Mehrotra, Manolis Zampetakis, Paul Kassianik, Blaine Nelson, Hyrum Anderson, Yaron Singer, and Amin Karbasi. Tree of attacks: Jailbreaking black-box llms automatically. *Advances in Neural Information Processing Systems*, 37:61065–61105, 2024.

Keivalya Pandya and Mehfuza Holia. Automating customer service using langchain: Building custom open-source gpt chatbot for organizations. *arXiv preprint arXiv:2310.05421*, 2023.

Harris Papadopoulos, Kostas Proedrou, Volodya Vovk, and Alex Gammerman. Inductive confidence machines for regression. In *European conference on machine learning*, pp. 345–356. Springer, 2002.

Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, pp. 1–22, 2023.

Sangdon Park, Osbert Bastani, Nikolai Matni, and Insup Lee. Pac confidence sets for deep neural networks via calibrated prediction. *arXiv preprint arXiv:2001.00106*, 2019.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019.

Coby Penso and Jacob Goldberger. A conformal prediction score that is robust to label noise. *arXiv preprint arXiv:2405.02648*, 2024.

Ethan Perez, Saffron Huang, Francis Song, Trevor Cai, Roman Ring, John Aslanides, Amelia Glaese, Nat McAleese, and Geoffrey Irving. Red teaming language models with language models. *arXiv preprint arXiv:2202.03286*, 2022.

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. SQuAD: 100,000+ questions for machine comprehension of text. In Jian Su, Kevin Duh, and Xavier Carreras (eds.), *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 2383–2392, Austin, Texas, November 2016. Association for Computational Linguistics.

Pranav Rajpurkar, Robin Jia, and Percy Liang. Know what you don't know: Unanswerable questions for SQuAD. In Iryna Gurevych and Yusuke Miyao (eds.), *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 784–789, Melbourne, Australia, July 2018. Association for Computational Linguistics.

Traian Rebedea, Razvan Dinu, Makesh Narsimhan Sreedhar, Christopher Parisien, and Jonathan Cohen. Nemo guardrails: A toolkit for controllable and safe llm applications with programmable rails. In *Proceedings of the 2023 conference on empirical methods in natural language processing: system demonstrations*, pp. 431–445, 2023.

Alexander Robey, Eric Wong, Hamed Hassani, and George J Pappas. Smoothllm: Defending large language models against jailbreaking attacks. *arXiv preprint arXiv:2310.03684*, 2023.

James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association*, 89(427): 846–866, 1994.

Matteo Sesia and Vladimir Svetnik. Doubly Robust Conformalized Survival Analysis with Right-Censored Data. *arXiv preprint arXiv:2412.09729*, 2024.

Matteo Sesia and Vladimir Svetnik. Conformal survival bands for risk screening under right-censoring. *arXiv preprint arXiv:2505.04568*, 2025.

Burr Settles. Active learning literature survey. *University of Wisconsin, Madison*, 52, 07 2010.

Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *Journal of machine learning research*, 9(3), 2008.

Gemma Team. Gemma 3 technical report. *arXiv preprint arXiv:2503.19786*, 2025.

Qwen Team. Qwen2.5: A party of foundation models, September 2024. URL `https://qwenlm.github.io/blog/qwen2.5/`.

Ryan J Tibshirani, Rina Foygel Barber, Emmanuel Candes, and Aaditya Ramdas. Conformal prediction under covariate shift. *Advances in neural information processing systems*, 32, 2019.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Bertie Vidgen, Nino Scherrer, Hannah Rose Kirk, Rebecca Qian, Anand Kannappan, Scott A Hale, and Paul Röttger. SimpleSafetyTests: a Test Suite for Identifying Critical Safety Risks in Large Language Models. *arXiv preprint arXiv:2311.08370*, 2023.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*, volume 29. Springer, 2005. doi: 10.1007/b106715.

Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge university press, 2019.

CY Wang and Hua Yun Chen. Augmented inverse probability weighted estimator for cox missing covariate regression. *Biometrics*, 57(2):414–419, 2001.

Ganghua Wang, Zhaorun Chen, Bo Li, and Haifeng Xu. Cer-eval: Certifiable and cost-efficient evaluation framework for llms. *arXiv preprint arXiv:2505.03814*, 2025.

Lei Wang, Chen Ma, Xueyang Feng, Zeyu Zhang, Hao Yang, Jingsen Zhang, Zhiyuan Chen, Jiakai Tang, Xu Chen, Yankai Lin, et al. A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6):186345, 2024a.

Liang Wang, Nan Yang, Xiaolong Huang, Linjun Yang, Rangan Majumder, and Furu Wei. Multilingual e5 text embeddings: A technical report. *arXiv preprint arXiv:2402.05672*, 2024b.

Yuxin Wang, Dennis Frauen, Jonas Schweisthal, Maresa Schröder, Emil Javurek, and Stefan Feuerriegel. Adaptive experimentation for censored survival outcomes. *arXiv preprint arXiv:2605.18459*, 2026a.

Yuyao Wang, Alexander W Levis, Shu Yang, and Larry Han. History-aware conformal prediction sets for censored time-to-event outcomes. *arXiv preprint arXiv:2605.06581*, 2026b.

Alexander Wei, Nika Haghtalab, and Jacob Steinhardt. Jailbroken: How does llm safety training fail? *Advances in neural information processing systems*, 36:80079–80110, 2023.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. Huggingface's transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*, 2019.

Skyler Wu, Yash Nair, and Emmanuel J Candès. Efficient evaluation of llm performance with statistical guarantees. *arXiv preprint arXiv:2601.20251*, 2026.

Yangxinyu Xie, Xuyang Chen, Zhimei Ren, and Weijie J Su. Watermark in the classroom: A conformal framework for adaptive ai usage detection. *arXiv preprint arXiv:2507.23113*, 2025.

Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, et al. Judging llm-as-a-judge with mt-bench and chatbot arena. *Advances in neural information processing systems*, 36:46595–46623, 2023.

Jin Peng Zhou, Christian K Belardi, Ruihan Wu, Travis Zhang, Carla P Gomes, Wen Sun, and Kilian Q Weinberger. On speeding up language model evaluation. In *The Thirteenth International Conference on Learning Representations*, 2025.

Shuyan Zhou, Frank F Xu, Hao Zhu, Xuhui Zhou, Robert Lo, Abishek Sridhar, Xianyi Cheng, Tianyue Ou, Yonatan Bisk, Daniel Fried, et al. Webarena: A realistic web environment for building autonomous agents. In *International Conference on Learning Representations*, volume 2024, pp. 15585–15606, 2024.

Andy Zou, Zifan Wang, Nicholas Carlini, Milad Nasr, J Zico Kolter, and Matt Fredrikson. Universal and transferable adversarial attacks on aligned language models. *arXiv preprint arXiv:2307.15043*, 2023.

Tijana Zrnic and Emmanuel Candès. Active Statistical Inference. In *International Conference on Machine Learning*, 2024.

A Additional related work

Our theoretical framework builds upon conformal prediction (Vovk et al., 2005; Shafer & Vovk, 2008), a powerful framework for constructing prediction sets with distribution-free finite-sample guarantees under exchangeability. Split conformal prediction (Papadopoulos et al., 2002) and its extension to covariate shift via weighting (Tibshirani et al., 2019), are utilized by the calibration procedure we adopt from (Davidov et al., 2026). Conformal prediction has recently been extended to survival analysis (Candès et al., 2023; Gui et al., 2024; Davidov et al., 2025), including history-aware prediction sets that use longitudinal covariates and inverse-probability-of-censoring weighting (Wang et al., 2026b). In contrast, we use the evolving history to determine which calibration trajectories to continue, while the final LPB remains a function of the initial prompt alone. Our work builds directly on Davidov et al. (2026), who apply conformalized survival analysis to LLM safety by introducing the time-to-unsafe-sampling metric and a static, variance-minimizing budget allocation scheme. We extend this framework to the multi-turn setting where conversation evolves, and our coverage proof provides a tighter PAC-type guarantee (Park et al., 2019; Bates et al., 2021; 2023).

The variance-minimization phase of our framework is related to active learning (Settles, 2010) and optimal experimental design (Fedorov, 2013), which study how to allocate a limited budget across observations to minimize estimator variance. In particular, Zrnic & Candès (2024) formalize active data collection for statistical inference while maintaining valid guarantees, and Hahn et al. (2011) develop adaptive designs that minimize treatment effect estimation variance. Recent work also studies adaptive experimentation for censored survival outcomes (Wang et al., 2026a); unlike its budget allocation across participants for causal survival effects, our method allocates budget within each trajectory to calibrate predictive bounds. Our contribution is adapting these ideas to the survival analysis and LLM evaluation setting, where observations arrive sequentially, and, in our framework, we learn an allocation policy from a first data split before being projected onto a second.

Prior efforts in LLM safety evaluation mostly focus on detecting harmful content in model outputs (Vidgen et al., 2023; Gehman et al., 2020) or constructing guardrails at inference time (Inan et al., 2023; Rebedea et al., 2023). A complementary line of work takes an adversarial perspective, generating red-teaming attacks to probe model vulnerabilities (Perez et al., 2022; Zou et al., 2023; Wei et al., 2023; Mehrotra et al., 2024; Chao et al., 2025). These works aim to maximize attack success. On the defensive side, recent efforts in certified LLM safety aim to provide statistical guarantees against such attacks, typically by bounding the failure probability of a single prompt using techniques like randomized smoothing (Kumar et al., 2023; Robey et al., 2023). In contrast, our goal is to provide rigorous statistical guarantees on time-to-event using the multi-turn interaction structure. While existing adversarial red-teaming approaches effectively investigate multi-turn vulnerabilities, they are not designed to estimate a censored time-to-event under resource constraints.

Our framework can be further leveraged for sample-efficient LLM evaluation (Kossen et al., 2021). To reduce computational costs and API calls when evaluating large models, recent methods utilize active testing (Li et al., 2024), multi-armed bandits (Zhou et al., 2025), adaptive dataset partitioning (Wang et al., 2025), and finite-population inference with adaptive benchmark-item selection (Wu et al., 2026). These methods select which fixed benchmark items to evaluate; in contrast, our framework allocates how far to roll out each sequential trajectory, inducing censoring that must be accounted for when constructing predictive bounds and estimating population-level metrics. By framing metric evaluation as a sequential, budget-constrained survival problem, `DAPRO` produces theoretically valid estimates with low variance under resource constraints.

B Extension to LLM utility evaluation and upper predictive bounds

While the main text focuses on constructing lower predictive bounds, our framework can be utilized for constructing upper predictive bounds, which are useful for evaluating LLM utility in multi-turn interactions, such as (Zhou et al., 2024). In this setting, the adversarial "attacker" is replaced by a "helper agent" that aims to guide the target model toward successful task completion. For example, in the AutoIF dataset (Dong et al., 2025), the helper agent iteratively assists the target model in satisfying complex instructions. The event of interest, $Y_i(t) = 1$, now represents successful task completion, and T_i denotes the corresponding time-to-success. We aim to construct an upper predictive bound

(UPB) $\widehat{U}(X)$ for this time-to-success: a smaller value indicates that success should occur within fewer interactions.

Finite-horizon target. In our framework, our interaction horizon is capped at $t_{\max}$. Since we do not observe events beyond time $t_{\max}$ and cannot guarantee whether a UPB that is greater than $t_{\max}$ correctly covers the time-to-success, we represent "no event by turn $t_{\max}$" by the additional value $t_{\max}+1$, which is treated as infinity, and define $T_i^{(t_{\max})} := \min\{T_i, t_{\max}+1\}$. Our desired, tractable guarantee in this UPB setting is therefore

$$\mathbb{P}\Big\{T_{\text{test}}^{(t_{\max})} \le \widehat{U}(X_{\text{test}}) \;\Big|\; \mathcal{D}_{\text{all}}\Big\} \ge 1-\alpha. \tag{7}$$

Thus, $t_{\max}$ is a finite bound that can fail to cover a trajectory with no event through turn $t_{\max}$, whereas $t_{\max}+1$ is an upper bound that always covers $T_{\text{test}}^{(t_{\max})}$. We remark that our acquisition is still truncated at $t_{\max}$, so no trajectory is acquired at time $t_{\max}+1$.

Candidate-specific UPB errors. For a candidate upper bound $\hat{f}_\tau(X_i)$, we define the error event as

$$A_i(\tau) := \mathbb{I}\{\hat{f}_\tau(X_i) < t_{\max}+1\}\mathbb{I}\{T_i > \hat{f}_\tau(X_i)\}. \tag{8}$$

Unlike a miscoverage event for an LPB, here an error is observed when either we observe an event or the acquisition policy reaches $\hat{f}_\tau(X_i)$. Consequently, the resolution time and reach probability depend on τ:

$$r_i(\tau) := \min\{T_i, \hat{f}_\tau(X_i)\}, \qquad \rho_i(t) := \prod_{s=1}^{t} P_i(s), \qquad \pi_i(\tau) := \rho_i(r_i(\tau)). \tag{9}$$

Above, $\rho_i(0) := 1$. Observe that the contribution of a sample i to the miscoverage estimator is $A_i(\tau)\mathbb{I}\{C_i \ge \hat{f}_\tau(X_i)\}/\pi_i(\tau)$; for τ having $\hat{f}_\tau(X_i) = t_{\max}+1$, the contribution is deterministically zero.

Sequential augmented miscoverage estimator. We construct the UPB using the outputs of the survival model at every reached prefix. For convenience, we write $H_i(0) = H_i(1)$. Specifically, for a candidate bound $f = \hat{f}_\tau(X_i)$, let

$$m_{i,0}(f) := \widehat{\mathbb{P}}(T_i > f \mid H_i(0)), \qquad m_{i,t}(f) := \widehat{\mathbb{P}}(T_i > f \mid H_i(t)) \quad (t < r_i(\tau)), \tag{10}$$

and set $m_{i,r_i(\tau)}(f) := A_i(\tau)$. We define $O_i(t) = \mathbb{I}\{\min(C_i, T_i) \ge t\}$ as a binary indicator that indicates that the turn t was acquired for the i-th sample. The miscoverage estimator of the i-th sample is

$$\widetilde{A}_i(\tau) := m_{i,0}(\hat{f}_\tau(X_i)) + \sum_{t=1}^{r_i(\tau)} \frac{O_i(t)}{\rho_i(t)} \Big[m_{i,t}(\hat{f}_\tau(X_i)) - m_{i,t-1}(\hat{f}_\tau(X_i))\Big]. \tag{11}$$

For $\hat{f}_\tau(X_i) = t_{\max}+1$, we instead set $\widetilde{A}_i(\tau) = 0$. We study the effect of the miscoverage estimator in Appendix F.4.9, analyzing either the standard estimator $\frac{1}{\pi_i(\tau)} O_i(r_i(\tau))A_i(\tau)$, or the above augmented sequential estimator. This study reveals that the augmented sequential estimator leads to a lower variance of the coverage rate.

Notice that equation 11 starts from the model's initial guess and adds an inverse-probability correction for every prediction change that the acquisition reveals. In expectation, these corrections reduce to the observed endpoint $A_i(\tau)$, so the accuracy of the model $m_{i,t}(f)$ affects variance but not validity. This form is inspired by prediction-powered inference (Angelopoulos et al., 2023), and based on the augmented Horvitz-Thompson (AHT) estimator (Horvitz & Thompson, 1952; Robins et al., 1994; Wang & Chen, 2001).

Proposition 1 (Validity of the UPB acquisition estimator)**.** *Conditionally on $\mathcal{D}_{\text{train}}, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}$, for every candidate τ,*

$$\mathbb{E}\Big[\widetilde{A}_i(\tau)\Big] = A_i(\tau). \tag{12}$$

This identity holds for any sequence of model predictions $m_{i,t}$. Moreover,

$$\text{Var}_{\text{acq}}\Big(\widetilde{A}_i(\tau) \mid X_i, H_i, T_i\Big) = \sum_{t=1}^{r_i(\tau)} \left(\frac{1}{\rho_i(t)} - \frac{1}{\rho_i(t-1)}\right) \Big(A_i(\tau) - m_{i,t-1}(\hat{f}_\tau(X_i))\Big)^2. \tag{13}$$

Proof. Fix i and τ, and condition throughout on $\mathcal{D}_{\text{train}}, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, X_i, H_i, T_i$, so that $r_i(\tau)$, $A_i(\tau)$, and the sequence $m_{i,t}(\hat{f}_\tau(X_i))$ are fixed. All expectations and variances below are therefore with respect to the acquisition randomness only.

Write

$$r := r_i(\tau), \qquad A := A_i(\tau), \qquad m_t := m_{i,t}(\hat{f}_\tau(X_i)), \qquad \rho_t := \rho_i(t), \qquad O_t := O_i(t)$$

and set $O_0 = \rho_0 = 1$. For $t \le r$, let $p_t := \rho_t/\rho_{t-1}$ denote the conditional probability of acquiring turn t, given that turn $t-1$ has been reached. We denote by $\mathcal{G}_t$ the information revealed by the acquisition process up to turn t.

$$\mathbb{E}[O_t \mid \mathcal{G}_{t-1}, X_i, H_i, T_i] = p_t O_{t-1},$$

and consequently

$$\mathbb{E}\left[\frac{O_t}{\rho_t} \,\middle|\, \mathcal{G}_{t-1}, X_i, H_i, T_i\right] = \frac{O_{t-1}}{\rho_{t-1}}.$$

In particular,

$$\mathbb{E}\left[\frac{O_t}{\rho_t}\right] = 1.$$

Hence, using $m_r = A$,

$$\begin{aligned}\mathbb{E}[\widetilde{A}_i(\tau)] &= m_0 + \sum_{t=1}^{r} \mathbb{E}\left[\frac{O_t}{\rho_t}\right](m_t - m_{t-1}) \\ &= m_0 + \sum_{t=1}^{r}(m_t - m_{t-1}) = m_r = A.\end{aligned}$$

This proves

$$\mathbb{E}\Big[\widetilde{A}_i(\tau) \mid X_i, H_i, T_i, \mathcal{D}_{\text{train}}, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}\Big] = A_i(\tau),$$

for an arbitrary sequence of predictions $m_{i,t}$.

We next establish the variance identity. Define

$$Q_t := \frac{O_t}{\rho_t}, \qquad Q_0 := 1, \qquad D_t := Q_t - Q_{t-1}.$$

The preceding conditional-expectation identity gives

$$\mathbb{E}[D_t \mid \mathcal{G}_{t-1}, X_i, H_i, T_i] = 0,$$

so $(D_t)_{t=1}^{r}$ is a martingale-difference sequence. Since $m_r = A$, a summation by parts yields

$$\begin{aligned}\widetilde{A}_i(\tau) - A &= m_0 - A + \sum_{t=1}^{r} Q_t(m_t - m_{t-1}) \\ &= \sum_{t=1}^{r} D_t\,(A - m_{t-1}).\end{aligned}$$

Since the martingale differences are pairwise orthogonal:

$$\mathrm{Var}(\widetilde{A}_i(\tau) \mid X_i, H_i, T_i) = \sum_{t=1}^{r} (A - m_{t-1})^2 \, \mathbb{E}[D_t^2].$$

Moreover, since $O_t O_{t-1} = O_t$,

$$\begin{aligned}\mathbb{E}[D_t^2] &= \mathbb{E}\left[\left(\frac{O_t}{\rho_t} - \frac{O_{t-1}}{\rho_{t-1}}\right)^2\right] \\ &= \frac{\mathbb{E}[O_t]}{\rho_t^2} + \frac{\mathbb{E}[O_{t-1}]}{\rho_{t-1}^2} - \frac{2\mathbb{E}[O_t]}{\rho_t \rho_{t-1}} \\ &= \frac{1}{\rho_t} - \frac{1}{\rho_{t-1}},\end{aligned}$$

because $\mathbb{E}[O_t] = \rho_t$. Therefore

$$\mathrm{Var}\Big(\widetilde{A}_i(\tau) \mid X_i, H_i, T_i\Big) = \sum_{t=1}^{r_i(\tau)} \left(\frac{1}{\rho_i(t)} - \frac{1}{\rho_i(t-1)}\right) \Big(A_i(\tau) - m_{i,t-1}(\hat{f}_\tau(X_i))\Big)^2.$$

For the candidate $\hat{f}_\tau(X_i) = t_{\max} + 1$, $A_i(\tau) = 0$ and $\widetilde{A}_i(\tau) = 0$ by definition, and hence its acquisition variance is identically zero. □

We estimate the miscoverage rate of a candidate τ by

$$\widehat{\alpha}_{\mathrm{UPB}}(\tau) := \frac{1}{|\mathcal{I}_{\mathrm{cal}}|} \sum_{i \in \mathcal{I}_{\mathrm{cal}}} \widetilde{A}_i(\tau). \tag{14}$$

As in the LPB construction, we restrict the candidate search space to $\mathcal{T} \subseteq [0, \tau_{\mathrm{prior}}]$. In this UPB case, a reasonable prior level for $1 - \alpha = 0.8$ is $\tau_{\mathrm{prior}} = 0.98$. Because UPB coverage is nondecreasing with the candidate level τ, we select

$$\widehat{\tau} := \inf\left\{\tau \le \tau_{\mathrm{prior}} : \sup_{\tau' \in [\tau, \tau_{\mathrm{prior}}]} \widehat{\alpha}_{\mathrm{UPB}}(\tau') \le \alpha\right\}. \tag{15}$$

The calibrated bound is $\widehat{U}(x) := \hat{f}_{\widehat{\tau}}(x)$. If no candidate up to τ_{prior} satisfies this rule, we return $\widehat{U}(x) = t_{\max} + 1$.

Theorem 3 (UPB coverage validity)**.** *Suppose the assumptions of Theorem 5 hold. In addition, suppose that $m_{i,t}(f) \in [0,1]$, $\rho_i(t) \ge \varepsilon > 0$ for every reachable prefix, and that*

$$\mathbb{E}\big[\pi_i(\tau)^{-1}\big] \le \bar{w}. \tag{16}$$

Let

$$\Delta_{\mathrm{UPB}} := \frac{\log(1/\delta)}{3|\mathcal{I}_{\mathrm{cal}}|\varepsilon} + \sqrt{\frac{\log^2(1/\delta)}{9|\mathcal{I}_{\mathrm{cal}}|^2\varepsilon^2} + \frac{2\log(1/\delta)(\bar{w} - \alpha^2)}{|\mathcal{I}_{\mathrm{cal}}|}}. \tag{17}$$

Then, with probability at least $1 - \delta$, the selector in equation 15 satisfies

$$\mathbb{P}\Big\{T_{\mathrm{test}}^{(t_{\max})} \le \widehat{U}(X_{\mathrm{test}}) \;\Big|\; \mathcal{D}_{\mathrm{all}}\Big\} \ge 1 - \alpha - \Delta_{\mathrm{UPB}}. \tag{18}$$

Proof. Let $p_\tau := \mathbb{E}[A_i(\tau)]$ and $Y_i := p_\tau - \widetilde{A}_i(\tau)$. The proof follows from the arguments in Theorem 5; with the change that we replace the variance of the miscoverage estimator with $\bar{w}$. For a realized last reached prefix k, summation by parts gives

$$\widetilde{A}_i = \frac{m_{i,k}}{\rho_i(k)} + \sum_{t=0}^{k-1} m_{i,t} \left(\frac{1}{\rho_i(t)} - \frac{1}{\rho_i(t+1)}\right), \tag{19}$$

so $1 - \varepsilon^{-1} \leq \widetilde{A}_i \leq \varepsilon^{-1}$. Proposition 1 also implies

$$\mathbb{E}[\widetilde{A}_i^2] = \mathbb{E}[\mathrm{Var}(\widetilde{A}_i) + A_i^2] \leq \mathbb{E}[\pi_i^{-1}] \leq \bar{w}. \tag{20}$$

We further replace the upper bound of $Y_i = P_\tau - \tilde{A}_i \leq \frac{1}{\varepsilon}$. The rest of the proof follows directly from the derivations in the proof of Theorem 5 by using the above bound on the variance of the miscoverage estimator. □

DAPRO for UPB construction. For constructing LPBs, Phase-I optimization uses the fixed reference level α, because the calibrated level $\widehat{\tau}$ is not yet known.

For constructing UPBs, since the calibration direction is reversed, we instead choose a reference level using the uncalibrated predictions of the survival model. Define

$$\widehat{a}_0(\tau) := \frac{1}{|\mathcal{I}_{\mathrm{cal1}}|} \sum_{i \in \mathcal{I}_{\mathrm{cal1}}} \mathbb{I}\{\hat{f}_\tau(X_i) \leq t_{\max}\} \widehat{\mathbb{P}}(T_i > \hat{f}_\tau(X_i) \mid H_i(0)), \tag{21}$$

Defining $\mathcal{T}_{\mathrm{prior}} = [0, \tau_{\mathrm{prior}}]$, the reference quantile level we use is

$$\tau_{\mathrm{ref}} := \min \left\{ \tau \in \mathcal{T}_{\mathrm{prior}} : \inf_{\tau' \in \mathcal{T}_{\mathrm{prior}},\, \tau' \geq \tau} \{1 - \widehat{a}_0(\tau')\} \geq 1 - \alpha \right\}. \tag{22}$$

If the set is empty, we set $\tau_{\mathrm{ref}} := \max \mathcal{T}_{\mathrm{prior}}$. We define $f^\star(X_i) := \hat{f}_{\tau_{\mathrm{ref}}}(X_i)$. Observe that this selection uses only the model outputs from the initial inputs for every trajectory, and does not require expending budget. This is the UPB analogue of fixing the LPB target at α, but it respects the reversed direction of upper-bound calibration and adapts automatically to the requested UPB coverage level.

Here, the score is the estimated probability for the event occurring at the current time $S_i(t) = \widehat{h}_i(t)$ defined in equation 134. Only the target coefficients change. Using the same global regularization δ as LPB DAPRO, the UPB coefficients are

$$\begin{aligned} a_i^{\mathrm{UPB}}(t) := \frac{1}{1+\delta} \Big[& \mathbb{I}\{t = f^\star(X_i),\ f^\star(X_i) \leq b_i,\ f^\star(X_i) \leq t_{\max}\} \big(1 - \widehat{h}_i(t)\big) \\ & + \delta\, \mathbb{I}\{t \leq b_i\} \widehat{h}_i(t) \Big]. \end{aligned} \tag{23}$$

For a specific reference bound $f^\star(X_i)$, $1 - \widehat{h}_i(f^\star(X_i))$ is the estimated conditional probability of surviving beyond this bound, given that the trajectory has reached it. The bound $t_{\max} + 1$ has deterministic zero miscoverage and does not contribute to the miscoverage estimator. We therefore replace the coefficients by $a_i^{\mathrm{UPB}}(t)$ and minimize the following objective:

$$\begin{aligned} \min_{M^{\mathrm{raw}}} \quad & \frac{1}{N_1} \sum_{i \in \mathcal{I}_{\mathrm{cal1}}} \sum_{t=1}^{b_i} \frac{a_i^{\mathrm{UPB}}(t)}{\prod_{s=1}^{t} M_s^{\mathrm{raw}}(S_i(s))} \\ \text{s.t.} \quad & \frac{1}{N_1} \sum_{i \in \mathcal{I}_{\mathrm{cal1}}} \mathcal{B}(M^{\mathrm{raw}}(S_i)) \leq \bar{B}_2, \\ & S_i(t) \leq S_j(t) \Longrightarrow M_t^{\mathrm{raw}}(S_i(t)) \leq M_t^{\mathrm{raw}}(S_j(t)) \quad \text{for every active } i, j, t. \end{aligned} \tag{24}$$

All remaining components, including the specific two-bin score-to-probability projection, optimization, cumulative-reach correction, probability positivity floor, and the optional CRC for budget control, are the same as in the LPB implementation in Section D.3.

C THEORY

In this section, we present our theoretical guarantees. For notational convenience, we slightly overload notation made in the first part and set $\mathcal{D}_{\mathrm{cal}} = \{(X_i, H_i, T_i)\}_{i \in \mathcal{I}_{\mathrm{cal}}}$, $\mathcal{D}_{\mathrm{cal1}} = \{(X_i, H_i, T_i)\}_{i \in \mathcal{I}_{\mathrm{cal1}}}$, $\mathcal{D}_{\mathrm{cal2}} = \{(X_i, H_i, T_i)\}_{i \in \mathcal{I}_{\mathrm{cal2}}}$, and $\mathcal{D}_{\mathrm{crc}} = \{(X_i, H_i, T_i)\}_{i \in \mathcal{I}_{\mathrm{crc}}}$. Hence, the combined data is denoted as $\mathcal{D}_{\mathrm{all}} = \mathcal{D}_{\mathrm{train}} \cup \mathcal{D}_{\mathrm{cal}}$.

Throughout this section, with a slight abuse of notation, $\mathcal{D}_{\text{all}}$ also includes all auxiliary algorithmic randomness realized during calibration and used to construct $\hat{L}$ (e.g., the continuation/censoring draws). Accordingly, conditioning on $\mathcal{D}_{\text{all}}$, and probabilities over its realization, include this randomness, so that $\hat{L}$ is fixed under the inner conditional probability.

C.1 Coverage rate guarantee of Algorithm 3

Before presenting the coverage validity guarantee of our proposed algorithm, we first show that the weights it uses are correct.

Proposition 2 (Valid weights of Algorithm 3)**.** *Suppose the continuation probabilities of Algorithm 3 satisfy $P_i(t) > 0$ almost surely for every $t \le \min(\hat{f}_{\text{prior}}(X_i), T_i)$ and $i \in \mathcal{I}_{\text{cal2}}$. For every $i \in \mathcal{I}_{\text{cal}}$ and every $\tau \in \mathcal{T} = [0, \tau_{\text{prior}}]$, the weights used by Algorithm 3 satisfy almost surely:*

$$\mathbb{E}\Big[w(i)\,\mathbb{I}\{T_i < \hat{f}_\tau(X_i)\}\,\mathbb{I}\{C_i \ge \hat{f}_\tau(X_i)\}\,\Big|\,X_i, H_i, T_i, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}\Big] = \mathbb{I}\{T_i < \hat{f}_\tau(X_i)\}. \tag{25}$$

Proof. We first consider $i \in \mathcal{I}_{\text{cal1}} \cup \mathcal{I}_{\text{crc}}$. By construction,

$$C_i = \hat{f}_{\text{prior}}(X_i) \qquad \text{and} \qquad w(i) = 1$$

deterministically. Since $\hat{f}_\tau$ is non-decreasing in τ and $\tau \le \tau_{\text{prior}}$, we have

$$\hat{f}_\tau(X_i) \le \hat{f}_{\text{prior}}(X_i) = C_i.$$

Consequently,

$$\mathbb{I}\{C_i \ge \hat{f}_\tau(X_i)\} = 1$$

deterministically, and therefore

$$\begin{aligned}\mathbb{E}\Big[w(i)\,\mathbb{I}\{T_i < \hat{f}_\tau(X_i)\}\,\mathbb{I}\{C_i \ge \hat{f}_\tau(X_i)\}\,\Big|\,X_i, H_i, T_i, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}\Big]\\ = \mathbb{I}\{T_i < \hat{f}_\tau(X_i)\}.\end{aligned}$$

Thus, equation 25 holds for every $i \in \mathcal{I}_{\text{cal1}} \cup \mathcal{I}_{\text{crc}}$. We now consider $i \in \mathcal{I}_{\text{cal2}}$. Recall that

$$P_i(t) = M_t(S_i(t))$$

is the continuation probability at time t, and that

$$\xi_i(t) = \xi_i(t-1) Z_t^{(i)}, \qquad Z_t^{(i)} \sim \text{Bern}(P_i(t)),$$

where the variables $\{Z_t^{(i)}\}_{t=1}^{t_{\max}}$ are mutually independent conditional on

$$X_i, H_i, T_i, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}.$$

Therefore, for every integer $k \ge 1$,

$$\begin{aligned}\mathbb{P}(\xi_i(k) = 1 \,|\, X_i, H_i, T_i, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}) &= \mathbb{P}\Big(Z_1^{(i)} = \cdots = Z_k^{(i)} = 1 \,\Big|\, X_i, H_i, T_i, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}\Big)\\ &= \prod_{t=1}^{k} P_i(t).\end{aligned} \tag{26}$$

Define

$$A_{i,\tau} := \mathbb{I}\{T_i < \hat{f}_\tau(X_i)\}$$

and

$$R_i := \xi_i\Big(\min(\hat{f}_{\text{prior}}(X_i), T_i)\Big)\,\mathbb{I}\{T_i < \hat{f}_{\text{prior}}(X_i)\}.$$

We first show that

$$\mathbb{I}\{T_i < \hat{f}_\tau(X_i)\}\mathbb{I}\{C_i \ge \hat{f}_\tau(X_i)\} = A_{i,\tau} R_i. \tag{27}$$

If $A_{i,\tau} = 0$, both sides of equation 27 are equal to zero. Suppose instead that $A_{i,\tau} = 1$. Since $\tau \le \tau_{\text{prior}}$ and $\hat{f}_\tau$ is non-decreasing in τ,

$$T_i < \hat{f}_\tau(X_i) \le \hat{f}_{\text{prior}}(X_i).$$

In particular,

$$T_i < \hat{f}_{\text{prior}}(X_i) \qquad \text{and} \qquad \min(\hat{f}_{\text{prior}}(X_i), T_i) = T_i,$$

so that

$$R_i = \xi_i(T_i).$$

If $R_i = 1$, then the algorithm reaches and observes the event at time T_i. By the definition of C_i, it then sets

$$C_i = \hat{f}_{\text{prior}}(X_i) \ge \hat{f}_\tau(X_i).$$

Conversely, if $R_i = 0$, then the algorithm halts before observing the event at time T_i. In this case,

$$C_i < T_i < \hat{f}_\tau(X_i),$$

and hence

$$\mathbb{I}\{C_i \ge \hat{f}_\tau(X_i)\} = 0.$$

Thus, on the event $A_{i,\tau} = 1$,

$$\mathbb{I}\{C_i \ge \hat{f}_\tau(X_i)\} = R_i,$$

which proves equation 27.

Next, observe that whenever $A_{i,\tau} R_i = 1$, the algorithm has continued successfully through time T_i. Hence, the cumulative continuation probability is

$$\prod_{t=1}^{\min(\hat{f}_{\text{prior}}(X_i), T_i)} P_i(t),$$

and the corresponding weight is

$$w(i) = \left(\prod_{t=1}^{\min(\hat{f}_{\text{prior}}(X_i), T_i)} P_i(t) \right)^{-1}.$$

Observe that the following equality holds almost surely:

$$w(i) A_{i,\tau} R_i = A_{i,\tau} R_i \left(\prod_{t=1}^{\min(\hat{f}_{\text{prior}}(X_i), T_i)} P_i(t) \right)^{-1}. \tag{28}$$

If $A_{i,\tau} R_i = 0$, both sides of equation 28 are zero. Since the acquisition horizon is finite, the cumulative product is therefore strictly positive, and the weight is well defined.

Moreover, using the definition of R_i and equation 26,

$$\begin{aligned}
&\mathbb{E}[R_i \mid X_i, H_i, T_i, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}] \\
&\quad = \mathbb{I}\{T_i < \hat{f}_{\text{prior}}(X_i)\} \mathbb{P}\Big(\xi_i\Big(\min(\hat{f}_{\text{prior}}(X_i), T_i) \Big) = 1 \,\Big|\, X_i, H_i, T_i, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}} \Big) \\
&\quad = \mathbb{I}\{T_i < \hat{f}_{\text{prior}}(X_i)\} \prod_{t=1}^{\min(\hat{f}_{\text{prior}}(X_i), T_i)} P_i(t).
\end{aligned} \tag{29}$$

Using equation 27, equation 28, and equation 29, we obtain

$$\begin{aligned}
&\mathbb{E}\Big[w(i)\, \mathbb{I}\{T_i < \hat{f}_\tau(X_i)\}\, \mathbb{I}\{C_i \ge \hat{f}_\tau(X_i)\} \,\Big|\, X_i, H_i, T_i, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}} \Big] \\
&\quad = \mathbb{E}[w(i) A_{i,\tau} R_i \mid X_i, H_i, T_i, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}] \\
&\quad = A_{i,\tau} \left(\prod_{t=1}^{\min(\hat{f}_{\text{prior}}(X_i), T_i)} P_i(t) \right)^{-1} \mathbb{E}[R_i \mid X_i, H_i, T_i, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}] \\
&\quad = A_{i,\tau} \left(\prod_{t=1}^{\min(\hat{f}_{\text{prior}}(X_i), T_i)} P_i(t) \right)^{-1} \mathbb{I}\{T_i < \hat{f}_{\text{prior}}(X_i)\} \prod_{t=1}^{\min(\hat{f}_{\text{prior}}(X_i), T_i)} P_i(t).
\end{aligned}$$

On the event $A_{i,\tau} = 1$, we have already established that

$$T_i < \hat{f}_{\text{prior}}(X_i) \qquad \text{and} \qquad \min(\hat{f}_{\text{prior}}(X_i), T_i) = T_i.$$

Hence,

$$\begin{aligned} A_{i,\tau} \left(\prod_{t=1}^{T_i} P_i(t) \right)^{-1} \mathbb{I}\{T_i < \hat{f}_{\text{prior}}(X_i)\} \prod_{t=1}^{\min(\hat{f}_{\text{prior}}(X_i), T_i)} P_i(t) \\ = A_{i,\tau} \left(\prod_{t=1}^{T_i} P_i(t) \right)^{-1} \prod_{t=1}^{T_i} P_i(t) \\ = A_{i,\tau}. \end{aligned}$$

If $A_{i,\tau} = 0$, the expression is also zero. Therefore, in all cases,

$$\mathbb{E}\Big[w(i)\, \mathbb{I}\{T_i < \hat{f}_\tau(X_i)\}\, \mathbb{I}\{C_i \geq \hat{f}_\tau(X_i)\} \,\Big|\, X_i, H_i, T_i, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}} \Big] = \mathbb{I}\{T_i < \hat{f}_\tau(X_i)\}.$$

□

We now turn to present the main coverage rate guarantee.

Theorem 4. *Fix a miscoverage level $\alpha \in (0,1)$ and a tolerance level $\delta \in (0,1)$. Suppose that, conditional on $\mathcal{D}_{\text{train}}$, $\{(X_i, T_i, H_i)\}_{i \in \mathcal{I}_{\text{cal}}}$ and $(X_{test}, T_{test}, H_{\text{test}})$ are drawn i.i.d. Let $\mathcal{T} = [0, \tau_{\text{prior}}]$, assume that $\{\hat{f}_\tau\}_{\tau \in \mathcal{T}}$ is measurable with respect to $\mathcal{D}_{\text{train}}$, and that $\hat{f}_\tau(x)$ is non-decreasing and continuous in τ with $\hat{f}_\tau(x) > 0$ for all x and $\tau > 0$. Define $\hat{f}_{\text{prior}} = \hat{f}_{\tau_{\text{prior}}}$. Suppose the continuation probabilities of Algorithm 3 satisfy $P_i(t) > 0$ almost surely for every $t \leq \min(\hat{f}_{\text{prior}}(X_i), T_i)$ and $i \in \mathcal{I}_{\text{cal2}}$. For $i \in \mathcal{I}_{\text{cal2}}$, define the latent inverse-probability weight*

$$w^*(i) := \left(\prod_{t=1}^{\min\{\hat{f}_{\text{prior}}(X_i), T_i\}} P_i(t) \right)^{-1}.$$

Further assume that there exists a constant $\bar{\nu} > \alpha^2$, which may depend on $\mathcal{D}_{\text{train}}$ but not on $\mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}$, such that almost surely over $\mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}$:

$$\mathbb{E}[w^*(i) \mathbb{I}\{T_i < \hat{f}_{\text{prior}}(X_i)\} \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}] \leq \bar{\nu} \quad \text{for all } i \in \mathcal{I}_{\text{cal2}}.$$

Define the coverage gap:

$$\Delta := \frac{\log(1/\delta)}{3|\mathcal{I}_{\text{cal}}|} + \sqrt{\frac{\log^2(1/\delta)}{9|\mathcal{I}_{\text{cal}}|^2} + \frac{2(\bar{\nu} - \alpha^2)\log(1/\delta)}{|\mathcal{I}_{\text{cal}}|}}.$$

Then, with probability at least $1 - \delta$ over the draw of $\mathcal{D}_{\text{all}} = (\mathcal{D}_{\text{train}}, \mathcal{D}_{\text{cal}})$, the lower predictive bound $\hat{L}(x)$ generated by Algorithm 3 satisfies:

$$\mathbb{P}\Big[T_{test} \geq \hat{L}(X_{test}) \,\Big|\, \mathcal{D}_{\text{all}} \Big] \;\geq\; 1 - \alpha - \Delta.$$

Proof. First, we note that, in practice, we do not compute the entire sequential continuation indicators $\xi_i(t)$. However, for the sake of this proof, indicators $\xi_i(t)$ can be viewed as an infinite sequence generated via $Z_t^i \sim \text{Ber}(M_t(S_i(t)))$, which is dependent only on the history H_i, features X_i, and calibration data $\mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}$. Thus, by the construction of the censoring times and following the i.i.d. assumption, we have

1. Conditional on $(\mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}})$, the acquisition policy $M_t(\cdot)$ is fixed. By the i.i.d. sampling assumption, the latent trajectories $\{H_i(t)\}_t$ for $i \in \mathcal{I}_{\text{cal2}}$ are mutually independent, and the Bernoulli continuation draws $\{Z_t^i\}_t$ are independent across samples and depend only on each sample's own trajectory. Therefore, $\{(X_i, H_i, T_i, C_i)\}_{i \in \mathcal{I}_{\text{cal2}}}$ are mutually independent conditional on $(\mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}})$.

2. The marginal law of (X_i, T_i, H_i) for $i \in \mathcal{I}_{\text{cal2}}$ is independent of $\mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}$ given $\mathcal{D}_{\text{train}}$, and $\{(X_i, T_i, H_i)\}_{i\in\mathcal{I}_{\text{cal2}}}$ are mutually independent given $(\mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}})$.

By Proposition 2, the weights used by the algorithm satisfy, almost surely,

$$\mathbb{E}\Big[w(i)\,\mathbb{I}\{T_i < \hat{f}_\tau(X_i)\}\,\mathbb{I}\{C_i \geq \hat{f}_\tau(X_i)\}\,\Big|\,X_i, H_i, T_i, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}\Big] = \mathbb{I}\{T_i < \hat{f}_\tau(X_i)\}, \tag{30}$$

for all $i \in \mathcal{I}_{\text{cal}}$ such that $\tau \in \mathcal{T} = [0, \tau_{\text{prior}}]$.

Because the continuation probabilities $P_i(t)$ are measurable with respect to $\sigma(X_i, H_i, T_i, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}})$, their finite product, and hence $w^*(i)$, is measurable with respect to the same sigma-algebra. Moreover, $P_i(t) > 0$ almost surely at every relevant step, so the finite product is strictly positive.

Therefore, all assumptions of Theorem 5 hold, and thus the LPB generated by Algorithm 3 is a valid PAC-type LPB with the desired bounds. □

We now present the general theorem for coverage validity guarantee.

Theorem 5 (General Coverage Validity)**.** *Fix a miscoverage level $\alpha \in (0,1)$ and a tolerance level $\delta \in (0,1)$. Let $\mathcal{D}_{\text{train}}$ be the training set, and let $\mathcal{I}_{\text{cal}} = \mathcal{I}_{\text{cal1}} \sqcup \mathcal{I}_{\text{cal2}} \sqcup \mathcal{I}_{\text{crc}}$ be a disjoint partition of the calibration set indexes, with $|\mathcal{I}_{\text{cal1}}| + |\mathcal{I}_{\text{cal2}}| + |\mathcal{I}_{\text{crc}}| = |\mathcal{I}_{\text{cal}}|$. Suppose that conditional on $\mathcal{D}_{\text{train}}$, $\{(X_i, T_i, H_i)\}_{i\in\mathcal{I}_{\text{cal}}}$ and $(X_{test}, T_{test}, H_{\text{test}})$ are drawn i.i.d., and that $\hat{f}_\tau(x)$ is non-decreasing and continuous in τ, with $\hat{f}_\tau(x) > 0$ for all x and $\tau > 0$. We assume $\{\hat{f}_\tau\}_{\tau\in\mathcal{T}}$ is measurable with respect to $\mathcal{D}_{\text{train}}$. Let $\mathcal{T} = [0, \tau_{\text{prior}}]$ and define $\hat{f}_{\text{prior}}(x) := \hat{f}_{\tau_{\text{prior}}}(x)$. For $i \in \mathcal{I}_{\text{cal1}} \cup \mathcal{I}_{\text{crc}}$, the weights are given by $w(i) = 1$, and $C_i = \hat{f}_{\text{prior}}(X_i)$. For every $i \in \mathcal{I}_{\text{cal2}}$ and every $\tau \in \mathcal{T}$, assume that the weights satisfy, almost surely,*

$$\mathbb{E}\Big[w(i)\,\mathbb{I}\{T_i < \hat{f}_\tau(X_i)\}\,\mathbb{I}\{C_i \geq \hat{f}_\tau(X_i)\}\,\Big|\,X_i, H_i, T_i, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}\Big] = \mathbb{I}\{T_i < \hat{f}_\tau(X_i)\}. \tag{31}$$

For every $i \in \mathcal{I}_{\text{cal2}}$, assume that $w(i) \in [1, \infty)$ is measurable with respect to $\sigma(X_i, H_i, T_i, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}})$.

Assume the following conditions hold:

1. ***(Mutual independence)** Conditional on $(\mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}})$, the observations $\{(X_i, H_i, T_i, C_i)\}_{i\in\mathcal{I}_{\text{cal2}}}$ are mutually independent.*
2. ***(Marginal independence of calibration data)** The marginal law of (X_i, T_i, H_i) for $i \in \mathcal{I}_{\text{cal2}}$ is independent of $\mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}$ given $\mathcal{D}_{\text{train}}$, and $\{(X_i, T_i, H_i)\}_{i\in\mathcal{I}_{\text{cal2}}}$ are mutually independent given $(\mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}})$.*
3. ***(Bounded weighted errors)** There exists a constant $\bar{\nu} > \alpha^2$, which may depend on $\mathcal{D}_{\text{train}}$ but not on $\mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}$, such that almost surely over $\mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}$:*

$$\mathbb{E}[w(i)\mathbb{I}\{T_i < \hat{f}_{\text{prior}}(X_i)\} \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}] \leq \bar{\nu} \quad \text{for all } i \in \mathcal{I}_{\text{cal2}}.$$

Define the estimated miscoverage rate and calibrated quantile level by:

$$\hat{\alpha}(\tau) := \frac{1}{|\mathcal{I}_{\text{cal}}|}\sum_{i\in\mathcal{I}_{\text{cal}}} w(i)\,\mathbb{I}\Big\{\tilde{T}_i < \hat{f}_\tau(X_i) \leq C_i\Big\}, \qquad \hat{\tau} := \sup\Big\{\tau \in \mathcal{T} \;:\; \sup_{\tau' \leq \tau} \hat{\alpha}(\tau') \leq \alpha\Big\}.$$

By convention, $0 \in \mathcal{T}$ and $\hat{f}_0(x) = 0$. Hence $\hat{\alpha}(0) = 0$, so $\hat{\tau}$ is well defined.

Define the coverage gap:

$$\Delta := \frac{\log(1/\delta)}{3|\mathcal{I}_{\text{cal}}|} + \sqrt{\frac{\log^2(1/\delta)}{9|\mathcal{I}_{\text{cal}}|^2} + \frac{2\big(\bar{\nu} - \alpha^2\big)\log(1/\delta)}{|\mathcal{I}_{\text{cal}}|}}.$$

Then, with probability at least $1 - \delta$ over the draw of $\mathcal{D}_{\text{all}} = (\mathcal{D}_{\text{train}}, \mathcal{D}_{\text{cal}})$, the lower predictive bound $\hat{L}(x) = \hat{f}_{\hat{\tau}}(x)$ satisfies:

$$\mathbb{P}\Big[T_{test} \geq \hat{L}(X_{test}) \,\Big|\, \mathcal{D}_{\text{all}}\Big] \;\geq\; 1 - \alpha - \Delta.$$

Proof. We define the oracle miscoverage level by:

$$\tau(\alpha+\Delta) = \sup\left\{\lambda \in \mathcal{T} : \mathbb{P}\left(T < \hat{f}_\lambda(X) \mid \mathcal{D}_{\text{train}}\right) \le \alpha + \Delta\right\}. \tag{32}$$

The above set is non-empty since $\tau = 0$ satisfies $\mathbb{P}(T < \hat{f}_0(X) \mid \mathcal{D}_{\text{train}}) = \mathbb{P}(T < 0 \mid \mathcal{D}_{\text{train}}) = 0 \le \alpha + \Delta$ by the convention $\hat{f}_0(x) = 0$.

We set $\tau_* := \tau(\alpha+\Delta)$. If $\tau_* = \tau_{\text{prior}}$, then $\hat{\tau} \le \tau_{\text{prior}} = \tau_*$ deterministically. Moreover, left continuity gives $\mathbb{P}\left(T < \hat{f}_{\tau_*}(X) \mid \mathcal{D}_{\text{train}}\right) \le \alpha + \Delta$. Hence, in what follows, assume $\tau_* < \tau_{\text{prior}}$.

Observe that it suffices to show $1-\delta \le \mathbb{P}(\hat{\tau} \le \tau(\alpha+\Delta) \mid \mathcal{D}_{\text{train}})$, since on the event $\{\hat{\tau} \le \tau(\alpha+\Delta)\}$, the monotonicity of $\hat{f}_\tau$ in τ and the left-continuity of $\tau \mapsto \mathbb{P}(T \ge \hat{f}_\tau(X) \mid \mathcal{D}_{\text{all}})$ give:

$$\mathbb{P}(T \ge \hat{f}_{\hat{\tau}}(X) \mid \mathcal{D}_{\text{all}}) \ge \mathbb{P}(T \ge \hat{f}_{\tau(\alpha+\Delta)}(X) \mid \mathcal{D}_{\text{all}}) \ge 1 - \alpha - \Delta. \tag{33}$$

It therefore remains to show $\mathbb{P}(\hat{\tau} \le \tau(\alpha+\Delta) \mid \mathcal{D}_{\text{train}}) \ge 1-\delta$. Fix $0 < \varepsilon < \tau_{\text{prior}} - \tau_*$ and set $\lambda := \tau(\alpha+\Delta) + \varepsilon$. We claim that $\{\hat{\alpha}(\lambda) > \alpha\} \subseteq \{\hat{\tau} \le \lambda\}$. Indeed, if $\hat{\alpha}(\lambda) > \alpha$, then for every $\tau \ge \lambda$, $\sup_{\tau' \le \tau} \hat{\alpha}(\tau') \ge \hat{\alpha}(\lambda) > \alpha$, so no $\tau \ge \lambda$ belongs to the set $\{\tau \in \mathcal{T} : \sup_{\tau' \le \tau} \hat{\alpha}(\tau') \le \alpha\}$, hence $\hat{\tau} = \sup\{\tau \in \mathcal{T} : \sup_{\tau' \le \tau} \hat{\alpha}(\tau') \le \alpha\} \le \lambda$. It therefore suffices to show:

$$\mathbb{P}(\hat{\alpha}(\lambda) \le \alpha \mid \mathcal{D}_{\text{train}}) \le \delta, \tag{34}$$

since this gives $\mathbb{P}(\hat{\tau} \le \lambda \mid \mathcal{D}_{\text{train}}) \ge 1-\delta$ for every $0 < \varepsilon < \tau_{\text{prior}} - \tau_*$. Taking $\varepsilon \to 0^+$ and applying continuity of the probability measure to the decreasing family of events $\{\hat{\tau} \le \tau(\alpha+\Delta)+\varepsilon\}$ (whose intersection equals $\{\hat{\tau} \le \tau(\alpha+\Delta)\}$) gives:

$$\mathbb{P}(\hat{\tau} \le \tau(\alpha+\Delta) \mid \mathcal{D}_{\text{train}}) = \lim_{\varepsilon \to 0^+} \mathbb{P}(\hat{\tau} \le \tau(\alpha+\Delta) + \varepsilon \mid \mathcal{D}_{\text{train}}) \ge 1-\delta, \tag{35}$$

as required. We define:

$$S_\lambda := \sum_{i \in \mathcal{I}_{\text{cal}}} w(i)\, \mathbb{I}\{T_i < \hat{f}_\lambda(X_i)\}\, \mathbb{I}\{C_i \ge \hat{f}_\lambda(X_i)\}, \tag{36}$$

$$S_\lambda^{(1)} := \sum_{i \in \mathcal{I}_{\text{cal1}} \cup \mathcal{I}_{\text{crc}}} w(i)\, \mathbb{I}\{T_i < \hat{f}_\lambda(X_i)\}\, \mathbb{I}\{C_i \ge \hat{f}_\lambda(X_i)\}, \tag{37}$$

$$S_\lambda^{(2)} := \sum_{i \in \mathcal{I}_{\text{cal2}}} w(i)\, \mathbb{I}\{T_i < \hat{f}_\lambda(X_i)\}\, \mathbb{I}\{C_i \ge \hat{f}_\lambda(X_i)\}, \tag{38}$$

so that $S_\lambda = S_\lambda^{(1)} + S_\lambda^{(2)}$ and $\hat{\alpha}(\lambda) = S_\lambda / |\mathcal{I}_{\text{cal}}|$, since $\mathbb{I}\{T_i < \hat{f}_\lambda(X_i) \le C_i\} = \mathbb{I}\{\tilde{T}_i < \hat{f}_\lambda(X_i) \le C_i\}$.

By construction, for $i \in \mathcal{I}_{\text{cal1}} \cup \mathcal{I}_{\text{crc}}$, $C_i = \hat{f}_{\text{prior}}(X_i)$ and $w(i) = 1$ deterministically. Since $\hat{f}_\lambda(X_i) \le \hat{f}_{\text{prior}}(X_i)$, $\mathbb{I}\{C_i \ge \hat{f}_\lambda(X_i)\} = 1$ deterministically. This gives us:

$$S_\lambda^{(1)} = \sum_{i \in \mathcal{I}_{\text{cal1}} \cup \mathcal{I}_{\text{crc}}} \mathbb{I}\{T_i < \hat{f}_\lambda(X_i)\} =: N_\lambda^{(1)}. \tag{39}$$

Lastly, we set $A := \bar{\nu} - \alpha^2$. Observe that

$$\begin{aligned}
\bar{\nu} &\ge \mathbb{E}[w(i)\mathbb{I}\{T_i < \hat{f}_{\text{prior}}(X_i)\} \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}] && (40)\\
&\ge \mathbb{E}[\mathbb{I}\{T_i < \hat{f}_\lambda(X_i)\} \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}] && (41)\\
&> \alpha, && (42)
\end{aligned}$$

and thus $A > 0$.

Step 1: Chernoff bound given $\mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}$. By the tower property:

$$\mathbb{P}(\hat{\alpha}(\lambda) \le \alpha \mid \mathcal{D}_{\text{train}}) = \mathbb{E}\left[\mathbb{P}(N_\lambda^{(1)} + S_\lambda^{(2)} \le |\mathcal{I}_{\text{cal}}|\alpha \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}) \,\middle|\, \mathcal{D}_{\text{train}}\right]. \tag{43}$$

Conditioning on $\mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}$ makes $N_\lambda^{(1)}$ a deterministic non-negative integer. Notice that $\bar{\nu}$ is also deterministic given $\mathcal{D}_{\text{train}}$ by its definition. Since $N_\lambda^{(1)}$ is deterministic given $\mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}$, Markov's inequality applied to the non-negative random variable $\exp(-t(N_\lambda^{(1)} + S_\lambda^{(2)} - |\mathcal{I}_{\text{cal}}|\alpha)) \ge 0$ gives, for any $t > 0$:

$$\mathbb{P}(N_\lambda^{(1)} + S_\lambda^{(2)} \le |\mathcal{I}_{\text{cal}}|\alpha \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}) \le e^{t|\mathcal{I}_{\text{cal}}|\alpha - tN_\lambda^{(1)}} \cdot \mathbb{E}\left[e^{-tS_\lambda^{(2)}} \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}\right]. \tag{44}$$

Step 2: Bounding $\mathbb{E}[e^{-tS_\lambda^{(2)}} \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}]$. For $i \in \mathcal{I}_{\text{cal2}}$, define $V_i := w(i)\,\mathbb{I}\{T_i < \hat{f}_\lambda(X_i)\}\,\mathbb{I}\{C_i \geq \hat{f}_\lambda(X_i)\}$, so $S_\lambda^{(2)} = \sum_{i \in \mathcal{I}_{\text{cal2}}} V_i$.

Mutual independence of $\{V_i\}_{i \in \mathcal{I}_{\text{cal2}}}$ given $(\mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}})$.

By our assumption, $\{(X_i, H_i, T_i, C_i)\}_{i \in \mathcal{I}_{\text{cal2}}}$ are mutually independent conditional on $(\mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}})$. Moreover, $w(i)$ is measurable with respect to $\sigma(X_i, H_i, T_i, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}})$. Therefore, V_i is a measurable function of the i-th observation and the conditioning data, and hence $\{V_i\}_{i \in \mathcal{I}_{\text{cal2}}}$ are mutually independent conditional on $(\mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}})$. *Mean of V_i.* By the tower property and equation 31,

$$\begin{aligned}
\mathbb{E}[V_i \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}] &= \mathbb{E}[\mathbb{E}[V_i \mid X_i, H_i, T_i, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}] \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}] \\
&= \mathbb{E}\Big[\mathbb{I}\{T_i < \hat{f}_\lambda(X_i)\} \,\Big|\, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}\Big] \\
&= \mathbb{P}\Big(T < \hat{f}_\lambda(X) \,\Big|\, \mathcal{D}_{\text{train}}\Big) =: P_\lambda, \qquad (45)
\end{aligned}$$

where the last equality follows from the i.i.d. assumption and the independence of the marginal law of (X_i, T_i, H_i) from $\mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}$ given $\mathcal{D}_{\text{train}}$.

Second moment and variance of V_i.

Define

$$A_{i,\lambda} := \mathbb{I}\{T_i < \hat{f}_\lambda(X_i)\},$$

Notice that:

$$V_i^2 = w^2(i) A_{i,\lambda} \mathbb{I}\{C_i \geq \hat{f}_\lambda(X_i)\} = w(i) V_i$$

Since $w(i)$ is fixed conditional on $(X_i, H_i, T_i, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}})$,

$$\begin{aligned}
&\mathbb{E}\big[V_i^2 \,\big|\, X_i, H_i, T_i, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}\big] \\
&\quad = \mathbb{E}[w(i) V_i \mid X_i, H_i, T_i, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}] \\
&\quad = w(i) \mathbb{E}[V_i \mid X_i, H_i, T_i, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}] \\
&\quad = w(i) A_{i,\lambda}, \qquad (46)
\end{aligned}$$

where the last step used equation 31.

Taking conditional expectation with respect to $(\mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}})$ and applying the bounded mean-weight assumption gives

$$\begin{aligned}
\mathbb{E}[V_i^2 \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}] &= \mathbb{E}[w(i) A_{i,\lambda} \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}] \qquad (47) \\
&\leq \mathbb{E}\big[w(i) A_{i,\tau_{\text{prior}}} \,\big|\, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}\big] \qquad (48) \\
&\leq \bar{\nu}. \qquad (49)
\end{aligned}$$

Consequently,

$$\begin{aligned}
\mathrm{Var}(V_i \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}) &= \mathbb{E}[V_i^2 \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}] - \mathbb{E}[V_i \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}]^2 \\
&\leq \bar{\nu} - P_\lambda^2 \\
&\leq \bar{\nu} - \alpha^2 =: A, \qquad (50)
\end{aligned}$$

where the last inequality follows from $P_\lambda > \alpha + \Delta > \alpha$.

One-sided Bernstein MGF bound for $S_\lambda^{(2)}$. Define mean-zero variables $Y_i^{(2)} := P_\lambda - V_i$ for $i \in \mathcal{I}_{\text{cal2}}$. Since $V_i \geq 0$, we have $Y_i^{(2)} \leq P_\lambda \leq 1$.

Since $Y_i^{(2)} := P_\lambda - V_i$ and $P_\lambda = \mathbb{E}[V_i \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}]$ is deterministic given $(\mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}})$, we have:

$$\begin{aligned}
\mathrm{Var}(Y_i^{(2)} \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}) &= \mathrm{Var}(P_\lambda - V_i \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}) \\
&= \mathrm{Var}(V_i \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}) \\
&= \mathbb{E}[V_i^2 \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}] - P_\lambda^2 \qquad (51) \\
&\leq \bar{\nu} - P_\lambda^2 \\
&\leq \bar{\nu} - \alpha^2 =: A
\end{aligned}$$

where the second equality uses the fact that adding or subtracting a constant does not change variance, and the last inequality follows from the definition of the supremum $\tau(\alpha+\Delta)$, and since $\lambda = \tau(\alpha+\Delta)+\varepsilon > \tau(\alpha+\Delta)$, we have $P_\lambda > \alpha+\Delta$.

We apply (Wainwright, 2019, Proposition 2.14) for $Y_i^{(2)}$ with $Y_i^{(2)} \le 1$ a.s., and

$$\mathbb{E}\left[\left(Y_i^{(2)}\right)^2 \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}\right] = \text{Var}\left[Y_i^{(2)} \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}\right] \le A$$

to get:

$$\mathbb{E}\left[\exp\left(t \sum_{i\in\mathcal{I}_{\text{cal2}}} Y_i^{(2)}\right) \,\middle|\, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}\right] \le \exp\left(\frac{t^2|\mathcal{I}_{\text{cal2}}|A/2}{1-t/3}\right), \quad 0<t<3. \tag{52}$$

Since $S_\lambda^{(2)} = |\mathcal{I}_{\text{cal2}}|P_\lambda - \sum_{i\in\mathcal{I}_{\text{cal2}}} Y_i^{(2)}$:

$$\mathbb{E}\left[e^{-tS_\lambda^{(2)}} \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}\right] \le \exp\left(-t|\mathcal{I}_{\text{cal2}}|P_\lambda + \frac{t^2|\mathcal{I}_{\text{cal2}}|A/2}{1-t/3}\right), \quad 0<t<3. \tag{53}$$

The right-hand side of equation 53 depends only on the deterministic quantities P_λ, A, $|\mathcal{I}_{\text{cal2}}|$, and t, and in particular does not depend on $\mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}$.

Step 3: Bounding $\mathbb{E}[e^{-tN_\lambda^{(1)}} \mid \mathcal{D}_{\text{train}}]$. Under our assumptions, $\{(X_i,T_i)\}_{i\in\mathcal{I}_{\text{cal1}}\cup\mathcal{I}_{\text{crc}}}$ are i.i.d. with marginal law independent of $\mathcal{D}_{\text{cal2}}$ given $\mathcal{D}_{\text{train}}$. Hence $Z_i^{(1)} := \mathbb{I}\{T_i < \hat{f}_\lambda(X_i)\} \sim \text{Bernoulli}(P_\lambda)$ are i.i.d. given $\mathcal{D}_{\text{train}}$, and $N_\lambda^{(1)} = \sum_{i\in\mathcal{I}_{\text{cal1}}\cup\mathcal{I}_{\text{crc}}} Z_i^{(1)}$.

Define mean-zero $U_i := P_\lambda - Z_i^{(1)}$ for $i \in \mathcal{I}_{\text{cal1}} \cup \mathcal{I}_{\text{crc}}$, so $\{U_i\}_{i\in\mathcal{I}_{\text{cal1}}\cup\mathcal{I}_{\text{crc}}}$ are i.i.d. given $\mathcal{D}_{\text{train}}$. Since $Z_i^{(1)} \ge 0$, $U_i \le P_\lambda \le 1$, so the one-sided bound $U_i \le 1$ holds.

Variance of U_i bounded by A. Let $j \in \mathcal{I}_{\text{cal2}}$, with $w(j) \ge 1$.

$$\begin{aligned}
P_\lambda &= \mathbb{E}[A_{j,\lambda} \,|\, \mathcal{D}_{\text{train}}] && (54)\\
&\le \mathbb{E}\left[A_{j,\tau_{\text{prior}}} \,\middle|\, \mathcal{D}_{\text{train}}\right] && (55)\\
&\le \mathbb{E}\left[w(j)A_{j,\tau_{\text{prior}}} \,\middle|\, \mathcal{D}_{\text{train}}\right] && (56)\\
&= \mathbb{E}\left[\mathbb{E}\left[w(j)A_{j,\tau_{\text{prior}}} \,\middle|\, \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}\right] \,\middle|\, \mathcal{D}_{\text{train}}\right] && (57)\\
&\le \bar{\nu}, && (58)
\end{aligned}$$

Since $\bar{\nu} \ge \alpha$, and $P_\lambda > \alpha$, we have

$$\text{Var}(U_i \mid \mathcal{D}_{\text{train}}) = P_\lambda(1-P_\lambda) = P_\lambda - P_\lambda^2 \le \bar{\nu} - P_\lambda^2 \le \bar{\nu} - \alpha^2 = A. \tag{59}$$

We apply (Wainwright, 2019, Proposition 2.14) with $b=1$ and $\nu = A$ to each i.i.d. U_i, summing over $|\mathcal{I}_{\text{cal1}} \cup \mathcal{I}_{\text{crc}}|$ terms, and writing $N_\lambda^{(1)} = |\mathcal{I}_{\text{cal1}} \cup \mathcal{I}_{\text{crc}}|P_\lambda - \sum_{i\in\mathcal{I}_{\text{cal1}}\cup\mathcal{I}_{\text{crc}}} U_i$:

$$\begin{aligned}
\mathbb{E}\left[e^{-tN_\lambda^{(1)}} \mid \mathcal{D}_{\text{train}}\right] &= e^{-t|\mathcal{I}_{\text{cal1}}\cup\mathcal{I}_{\text{crc}}|P_\lambda} \cdot \mathbb{E}\left[\exp\left(t \sum_{i\in\mathcal{I}_{\text{cal1}}\cup\mathcal{I}_{\text{crc}}} U_i\right) \,\middle|\, \mathcal{D}_{\text{train}}\right] \\
&\le \exp\left(-t|\mathcal{I}_{\text{cal1}} \cup \mathcal{I}_{\text{crc}}|P_\lambda + \frac{t^2|\mathcal{I}_{\text{cal1}} \cup \mathcal{I}_{\text{crc}}|A/2}{1-t/3}\right), \quad 0<t<3.
\end{aligned} \tag{60}$$

Step 4: Combining the bounds. Applying equation 44 pointwise in $\mathcal{I}_{\text{cal1}} \cup \mathcal{I}_{\text{crc}}$, then taking the outer expectation over $\mathcal{I}_{\text{cal1}} \cup \mathcal{I}_{\text{crc}}$ via equation 43:

$$\mathbb{P}(\hat{\alpha}(\lambda) \le \alpha \mid \mathcal{D}_{\text{train}}) \le e^{t|\mathcal{I}_{\text{cal}}|\alpha} \cdot \mathbb{E}\left[e^{-tN_\lambda^{(1)}} \cdot \mathbb{E}\left[e^{-tS_\lambda^{(2)}} \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}\right] \,\middle|\, \mathcal{D}_{\text{train}}\right]. \tag{61}$$

Since the upper bound in equation 53 is deterministic (does not depend on $\mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}$), it factors out of the outer expectation:

$$\mathbb{P}(\hat{\alpha}(\lambda) \le \alpha \mid \mathcal{D}_{\text{train}}) \le e^{t|\mathcal{I}_{\text{cal}}|\alpha} \cdot \exp\left(-t|\mathcal{I}_{\text{cal2}}|P_\lambda + \frac{t^2|\mathcal{I}_{\text{cal2}}|A/2}{1-t/3}\right) \cdot \mathbb{E}\left[e^{-tN_\lambda^{(1)}} \mid \mathcal{D}_{\text{train}}\right]. \tag{62}$$

Applying equation 60 and using $|\mathcal{I}_{\text{cal1}}| + |\mathcal{I}_{\text{crc}}| + |\mathcal{I}_{\text{cal2}}| = |\mathcal{I}_{\text{cal}}|$:

$$\begin{aligned}\mathbb{P}(\hat{\alpha}(\lambda) \le \alpha \mid \mathcal{D}_{\text{train}}) &\le \exp\left(t|\mathcal{I}_{\text{cal}}|\alpha - t|\mathcal{I}_{\text{cal2}}|P_\lambda - t|\mathcal{I}_{\text{cal1}} \cup \mathcal{I}_{\text{crc}}|P_\lambda + \frac{t^2(|\mathcal{I}_{\text{cal2}}| + |\mathcal{I}_{\text{cal1}}| + |\mathcal{I}_{\text{crc}}|)A/2}{1 - t/3}\right) \\ &= \exp\left(-t|\mathcal{I}_{\text{cal}}|(P_\lambda - \alpha) + \frac{t^2|\mathcal{I}_{\text{cal}}|A/2}{1 - t/3}\right) \\ &\le \exp\left(-t|\mathcal{I}_{\text{cal}}|\Delta + \frac{t^2|\mathcal{I}_{\text{cal}}|A/2}{1 - t/3}\right) =: \exp(\mathcal{E}(t)),\end{aligned} \tag{63}$$

for $0 < t < 3$, where the last inequality uses $P_\lambda - \alpha \ge \Delta$ (since $\lambda > \tau(\alpha + \Delta)$ implies $P_\lambda = \mathbb{P}(T < \hat{f}_\lambda(X) \mid \mathcal{D}_{\text{train}}) \ge \alpha + \Delta$, by definition of $\tau(\alpha + \Delta)$ and continuity of $\hat{f}_\tau$).

Step 5: Evaluation at the Bernstein point and solving for Δ. Set $N := |\mathcal{I}_{\text{cal}}|$, and recall $A = \bar{\nu} - \alpha^2 > 0$. We evaluate $\mathcal{E}$ at:

$$t^* := \frac{\Delta}{A + \Delta/3}. \tag{64}$$

Notice that $t^* \in (0, 3)$ since $\Delta, A > 0$. Substituting into $\mathcal{E}(t^*)$:

$$\begin{aligned}\mathcal{E}(t^*) &= -\frac{N\Delta^2}{A + \Delta/3} + \frac{NA}{2} \cdot \frac{\Delta^2}{(A + \Delta/3)^2} \cdot \frac{1}{A/(A + \Delta/3)} \\ &= -\frac{N\Delta^2}{A + \Delta/3} + \frac{NA}{2} \cdot \frac{\Delta^2}{A(A + \Delta/3)} \\ &= -\frac{N\Delta^2}{A + \Delta/3} + \frac{N\Delta^2}{2(A + \Delta/3)} \\ &= -\frac{N\Delta^2}{2(A + \Delta/3)}\end{aligned} \tag{65}$$

Therefore,

$$\exp(\mathcal{E}(t^*)) \le \delta \iff \frac{N\Delta^2}{2(A + \Delta/3)} \ge \log\left(\frac{1}{\delta}\right). \tag{66}$$

Multiplying both sides by $2(A + \Delta/3) > 0$ and rearranging:

$$|\mathcal{I}_{\text{cal}}|\Delta^2 - \frac{2}{3}\log(1/\delta)\,\Delta - 2(\bar{\nu} - \alpha^2)\log\left(\frac{1}{\delta}\right) = 0. \tag{67}$$

This quadratic in Δ has positive leading coefficient $|\mathcal{I}_{\text{cal}}| > 0$ and negative constant term $-2A\log(1/\delta) < 0$. By Descartes' rule of signs, it has exactly one positive root:

$$\Delta := \frac{\log(1/\delta)}{3|\mathcal{I}_{\text{cal}}|} + \sqrt{\frac{\log^2(1/\delta)}{9|\mathcal{I}_{\text{cal}}|^2} + \frac{2(\bar{\nu} - \alpha^2)\log(1/\delta)}{|\mathcal{I}_{\text{cal}}|}}. \tag{68}$$

For the above Δ, condition equation 66 holds with equality, so $\mathbb{P}(\hat{\alpha}(\lambda) \le \alpha \mid \mathcal{D}_{\text{train}}) \le \exp(\mathcal{E}(t^*)) \le \delta$, satisfying equation 34.

Step 6: Conclusion. Since equation 34 holds for every $0 < \varepsilon < \tau_{\text{prior}} - \tau_*$ with $\lambda = \tau(\alpha + \Delta) + \varepsilon$, we have $\mathbb{P}(\hat{\tau} \le \lambda \mid \mathcal{D}_{\text{train}}) \ge 1 - \delta$. Taking $\varepsilon \to 0^+$ and applying continuity of the probability measure to the decreasing family of events $\{\hat{\tau} \le \tau(\alpha + \Delta) + \varepsilon\}$, whose intersection is $\{\hat{\tau} \le \tau(\alpha + \Delta)\}$:

$$\mathbb{P}(\hat{\tau} \le \tau(\alpha + \Delta) \,|\, \mathcal{D}_{\text{train}}) = \lim_{\varepsilon \to 0^+} \mathbb{P}(\hat{\tau} \le \tau(\alpha + \Delta) + \varepsilon \,|\, \mathcal{D}_{\text{train}}) \ge 1 - \delta. \tag{69}$$

On the event $\{\hat{\tau} \le \tau(\alpha + \Delta)\}$, monotonicity of $\hat{f}_\tau$ in τ gives $\hat{f}_{\hat{\tau}}(x) \le \hat{f}_{\tau(\alpha+\Delta)}(x)$ for all x, hence $\{T \ge \hat{f}_{\tau(\alpha+\Delta)}(X)\} \subseteq \{T \ge \hat{f}_{\hat{\tau}}(X)\}$. Since $\hat{f}_{\tau(\alpha+\Delta)}$ is $\mathcal{D}_{\text{train}}$-measurable and $(X_{\text{test}}, T_{\text{test}})$ is independent of $(\mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{cal2}})$ given $\mathcal{D}_{\text{train}}$:

$$\mathbb{P}\left[T_{\text{test}} \ge \hat{f}_{\hat{\tau}}(X_{\text{test}}) \,\middle|\, \mathcal{D}_{\text{all}}\right] \ge \mathbb{P}\left[T_{\text{test}} \ge \hat{f}_{\tau(\alpha+\Delta)}(X_{\text{test}}) \,\middle|\, \mathcal{D}_{\text{all}}\right] \ge 1 - \alpha - \Delta, \tag{70}$$

where the last inequality uses the definition of $\tau(\alpha+\Delta)$ and left-continuity of $\tau \mapsto \mathbb{P}(T < \hat{f}_\tau(X) \mid \mathcal{D}_{\text{train}})$, which give $\mathbb{P}(T < \hat{f}_{\tau(\alpha+\Delta)}(X) \mid \mathcal{D}_{\text{all}}) \leq \alpha+\Delta$. Notice that $\mathbb{P}(T < \hat{f}_{\tau(\alpha+\Delta)}(X) \mid \mathcal{D}_{\text{all}}) = \mathbb{P}(T < \hat{f}_{\tau(\alpha+\Delta)}(X) \mid \mathcal{D}_{\text{train}})$ since $\hat{f}_{\tau(\alpha+\Delta)}$ is $\mathcal{D}_{\text{train}}$-measurable and $(X_{\text{test}}, T_{\text{test}}) \perp\!\!\!\perp \mathcal{D}_{\text{cal}} \mid \mathcal{D}_{\text{train}}$. This holds with probability at least $1-\delta$ over $\mathcal{D}_{\text{cal}}$ given $\mathcal{D}_{\text{train}}$, hence over $\mathcal{D}_{\text{all}}$, completing the proof. □

Remark 1 (The assumption on $\bar{\nu}$ is enforced by construction)**.** *The implementation of* DAPRO *in Appendix D.3 floors the cumulative reach probability at $\varepsilon > 0$. Thus, the assumption that the mean weights of samples satisfying $T_i < \hat{f}_{\text{prior}}(X_i)$ holds by construction:*

$$\mathbb{E}[w(i)\mathbb{I}\{T_i < \hat{f}_{\text{prior}}(X_i)\} \mid \mathcal{D}_{\text{cal1}}, \mathcal{D}_{\text{crc}}, \mathcal{D}_{\text{train}}] \leq \bar{\nu} \quad \textit{for all } i \in \mathcal{I}_{\text{cal2}},$$

for $\bar{\nu} \leq \varepsilon^{-1}$.

Remark 2 (Discrete predictors and discrete search spaces)**.** *Theorems 4 and 5 hold when $\tau \mapsto \hat{f}_\tau(x)$ is discrete-valued and non-decreasing (rather than continuous), and the search space $\mathcal{T}$ is a finite grid. This is the setting of our experiments. Continuity is used in the proof only to guarantee that the supremum defining $\tau_* = \tau(\alpha+\Delta)$ is attained and to justify the limit $\varepsilon \to 0^+$; both are immediate when $\mathcal{T}$ is finite. Writing $P_\nu := \mathbb{P}(T < \hat{f}_\nu(X) \mid \mathcal{D}_{\text{train}})$, with the following changes, where Steps 1–5 are unchanged:*

1. *The definition of τ_* turns to $\tau_* := \max\{\nu \in \mathcal{T} : P_\nu \leq \alpha+\Delta\}$.*
2. *$\lambda = \tau_* + \varepsilon$ is replaced by $\lambda := \min\{t \in \mathcal{T} : t > \tau_*\}$*
3. *Since $\hat{\tau} \in \mathcal{T}$ and λ is the successor of τ_*, the inclusion $\{\hat{\alpha}(\lambda) > \alpha\} \subseteq \{\hat{\tau} < \lambda\}$ already gives $\hat{\tau} \leq \tau_*$. The limiting argument $\varepsilon \to 0^+$ in Step 6 is therefore not needed, and the conclusion follows from monotonicity of $\hat{f}_\tau$ together with $P_{\tau_*} \leq \alpha+\Delta$.*

The same substitutions apply to UPB coverage guarantee in Theorem 3, with the maximum replaced by a minimum and the successor of τ_* by the predecessor of $\tau^\dagger := \min\{t \in \mathcal{T} : \mathbb{P}(T^{(t_{\max})} > \hat{f}_t(X) \mid \mathcal{D}_{\text{train}}) \leq \alpha + \Delta_{\text{UPB}}\}$ in $\mathcal{T}$.

C.2 Budget validity of DAPRO via conformal risk control

In this section, show that DAPRO (with CRC), as outlined in Algorithm 3, satisfies the expected budget constraint in equation 1. Notice that to prove this we must assume that the budget is sufficiently large so that the set in equation 6 is not empty. Formally, this means that the total budget must satisfy almost surely:

$$B \geq \sum_{i \in \mathcal{I}_{\text{cal1}}} b_i + \frac{|\mathcal{I}_{\text{cal2}}| \sum_{i \in \mathcal{I}_{\text{crc}}} \mathcal{B}(M(S_i;0)) + |\mathcal{I}_{\text{crc}}| \sum_{i \in \mathcal{I}_{\text{crc}}} b_i + |\mathcal{I}_{\text{cal2}}| C_{\max} + |\mathcal{I}_{\text{crc}}| t_{\max}}{|\mathcal{I}_{\text{crc}}| + 1}.$$

Since the consumed budget is non-decreasing in λ, the above is exactly the condition that $\lambda = 0$ is feasible in equation 6, and hence guarantees that its feasible set is nonempty.

Proof of Theorem 2. By the construction of Algorithm 3, $\{(X_i, T_i, H_i)\}_{i \in \mathcal{I}_{\text{cal}}}$ are drawn exchangeably, independently of $\mathcal{D}_{\text{train}}$, and the data split into $\mathcal{I}_{\text{cal1}}, \mathcal{I}_{\text{crc}}$, and $\mathcal{I}_{\text{cal2}}$ is independent of the trajectories. Furthermore, conditional on $\mathcal{D}_{\text{train}}$ and the data used in Phase I, the candidate family $\{M_t(\cdot;\lambda)\}_{\lambda \in [0,1]}$ is fixed before observing $\mathcal{I}_{\text{crc}}$, the function $\mathcal{B}(M(S_i;\lambda))$ is continuous and non-decreasing in λ for every complete trajectory, and

$$0 \leq \mathcal{B}(M(S_i;\lambda)) \leq C_{\max}$$

for every candidate λ and every trajectory in the population support. Finally, the feasible set in equation 6 is nonempty by construction.

Recall that the samples in $\mathcal{I}_{\text{cal1}}$ and $\mathcal{I}_{\text{crc}}$ are fully observed until either the event occurs or $\hat{f}_{\text{prior}}(X_i)$ is reached, so that their acquisition cost is

$$b_i = \min\{T_i, \hat{f}_{\text{prior}}(X_i)\}.$$

Since $\hat{f}_{\mathrm{prior}}(X_i) \le t_{\max}$, we have $0 \le b_i \le t_{\max}$ a.s.

Condition on $\mathcal{D}_{\mathrm{train}}, \mathcal{D}_{\mathrm{cal1}}$ and on all information used to fit the acquisition policy in Phase I. Under this conditioning, the Phase-I budget consumption $\sum_{i\in\mathcal{I}_{\mathrm{cal1}}} b_i$, the remaining budget $B_{\mathrm{rem}} = B - \sum_{i\in\mathcal{I}_{\mathrm{cal1}}} b_i$, the candidate family $\{M_t(\cdot;\lambda)\}_{\lambda\in[0,1]}$, and $C_{\max}$ are fixed. Moreover, because the calibration trajectories are exchangeable and the split is independent of the trajectories, the complete trajectories in $\mathcal{I}_{\mathrm{crc}}$ and $\mathcal{I}_{\mathrm{cal2}}$ remain identically distributed and independent under this conditioning.

Consider an additional complete trajectory $(X_\star, T_\star, H_\star)$ drawn independently from the same distribution as the trajectories in $\mathcal{I}_{\mathrm{crc}}$ and $\mathcal{I}_{\mathrm{cal2}}$, and let

$$b_\star := \min\{T_\star, \hat{f}_{\mathrm{prior}}(X_\star)\}.$$

Let $S_\star$ denote the corresponding sequence of scores. We use this additional trajectory only for the proof.

Define the augmented CRC selector

$$\widetilde{\lambda} := \sup\left\{ \lambda \in [0,1] : \frac{\displaystyle\sum_{i\in\mathcal{I}_{\mathrm{crc}}} \left(\mathcal{B}(M(S_i;\lambda)) + \frac{|\mathcal{I}_{\mathrm{crc}}|}{|\mathcal{I}_{\mathrm{cal2}}|} b_i\right) + \mathcal{B}(M(S_\star;\lambda)) + \frac{|\mathcal{I}_{\mathrm{crc}}|}{|\mathcal{I}_{\mathrm{cal2}}|} b_\star}{|\mathcal{I}_{\mathrm{crc}}| + 1} \le \frac{B_{\mathrm{rem}}}{|\mathcal{I}_{\mathrm{cal2}}|} \right\}. \tag{71}$$

By assumption, $\mathcal{B}(M(S;\lambda))$ is continuous in λ. Hence the feasible sets in equation 6 and equation 71 are closed, and whenever they are nonempty, their suprema are themselves feasible.

By the row-wise cap and the bound $b_\star \le t_{\max}$,

$$\mathcal{B}(M(S_\star;\lambda)) + \frac{|\mathcal{I}_{\mathrm{crc}}|}{|\mathcal{I}_{\mathrm{cal2}}|} b_\star \le C_{\max} + \frac{|\mathcal{I}_{\mathrm{crc}}|}{|\mathcal{I}_{\mathrm{cal2}}|} t_{\max} \qquad \text{for every } \lambda \in [0,1]. \tag{72}$$

Using the definition of $\widehat{\lambda}$ in equation 6 together with equation 72, we therefore have

$$\begin{aligned}
&\sum_{i\in\mathcal{I}_{\mathrm{crc}}} \left(\mathcal{B}(M(S_i;\widehat{\lambda})) + \frac{|\mathcal{I}_{\mathrm{crc}}|}{|\mathcal{I}_{\mathrm{cal2}}|} b_i\right) + \mathcal{B}(M(S_\star;\widehat{\lambda})) + \frac{|\mathcal{I}_{\mathrm{crc}}|}{|\mathcal{I}_{\mathrm{cal2}}|} b_\star \\
&\quad \le \sum_{i\in\mathcal{I}_{\mathrm{crc}}} \left(\mathcal{B}(M(S_i;\widehat{\lambda})) + \frac{|\mathcal{I}_{\mathrm{crc}}|}{|\mathcal{I}_{\mathrm{cal2}}|} b_i\right) + C_{\max} + \frac{|\mathcal{I}_{\mathrm{crc}}|}{|\mathcal{I}_{\mathrm{cal2}}|} t_{\max} \\
&\quad \le (|\mathcal{I}_{\mathrm{crc}}| + 1) \frac{B_{\mathrm{rem}}}{|\mathcal{I}_{\mathrm{cal2}}|}.
\end{aligned} \tag{73}$$

Thus, $\widehat{\lambda}$ is feasible for the augmented problem in equation 71, and consequently

$$\widetilde{\lambda} \ge \widehat{\lambda}.$$

Because the candidate acquisition cost is non-decreasing in λ,

$$\mathcal{B}(M(S_\star;\widehat{\lambda})) + \frac{|\mathcal{I}_{\mathrm{crc}}|}{|\mathcal{I}_{\mathrm{cal2}}|} b_\star \le \mathcal{B}(M(S_\star;\widetilde{\lambda})) + \frac{|\mathcal{I}_{\mathrm{crc}}|}{|\mathcal{I}_{\mathrm{cal2}}|} b_\star. \tag{74}$$

The augmented selector $\widetilde{\lambda}$ is a permutation-invariant function of the $|\mathcal{I}_{\mathrm{crc}}| + 1$ exchangeable complete trajectories consisting of the CRC trajectories together with the additional trajectory $(X_\star, T_\star, H_\star)$. Therefore, by exchangeability,

$$\begin{aligned}
&\mathbb{E}\left[\mathcal{B}(M(S_\star;\widetilde{\lambda})) + \frac{|\mathcal{I}_{\mathrm{crc}}|}{|\mathcal{I}_{\mathrm{cal2}}|} b_\star\right] \\
&\quad = \mathbb{E}\left[\frac{1}{|\mathcal{I}_{\mathrm{crc}}| + 1} \left\{ \sum_{i\in\mathcal{I}_{\mathrm{crc}}} \left(\mathcal{B}(M(S_i;\widetilde{\lambda})) + \frac{|\mathcal{I}_{\mathrm{crc}}|}{|\mathcal{I}_{\mathrm{cal2}}|} b_i\right) + \mathcal{B}(M(S_\star;\widetilde{\lambda})) + \frac{|\mathcal{I}_{\mathrm{crc}}|}{|\mathcal{I}_{\mathrm{cal2}}|} b_\star \right\}\right] \\
&\quad \le \frac{B_{\mathrm{rem}}}{|\mathcal{I}_{\mathrm{cal2}}|},
\end{aligned} \tag{75}$$

where the final inequality follows from the feasibility of $\widetilde{\lambda}$ in equation 71. Combining equation 74 and equation 75 yields

$$\mathbb{E}\Big[\mathcal{B}(M(S_\star;\widehat{\lambda}))\Big] + \frac{|\mathcal{I}_{\mathrm{crc}}|}{|\mathcal{I}_{\mathrm{cal2}}|}\mathbb{E}[b_\star] \le \frac{B_{\mathrm{rem}}}{|\mathcal{I}_{\mathrm{cal2}}|}. \tag{76}$$

Multiplying both sides by $|\mathcal{I}_{\mathrm{cal2}}|$ gives

$$|\mathcal{I}_{\mathrm{cal2}}|\,\mathbb{E}\Big[\mathcal{B}(M(S_\star;\widehat{\lambda}))\Big] + |\mathcal{I}_{\mathrm{crc}}|\,\mathbb{E}[b_\star] \le B_{\mathrm{rem}}. \tag{77}$$

It remains to show that the two terms on the left-hand side of equation 77 are exactly the expected deployment and CRC expenditures.

First, every sample $i \in \mathcal{I}_{\mathrm{crc}}$ is fully observed until $b_i = \min\{T_i, \hat{f}_{\mathrm{prior}}(X_i)\}$. Since the CRC trajectories and the fresh trajectory $H_\star$ have the same distribution,

$$\mathbb{E}\left[\sum_{i\in\mathcal{I}_{\mathrm{crc}}} b_i\right] = |\mathcal{I}_{\mathrm{crc}}|\,\mathbb{E}[b_\star]. \tag{78}$$

Second, consider a deployment sample $i \in \mathcal{I}_{\mathrm{cal2}}$. Conditional on its complete trajectory and on $\widehat{\lambda}$, step t is acquired if and only if all continuation decisions through step t succeed. By the randomization in Algorithm 3,

$$\mathbb{P}\Big(\xi_i(t) = 1 \,\Big|\, X_i, T_i, H_i, \widehat{\lambda}\Big) = \prod_{j=1}^{t} M_j(S_i(j);\widehat{\lambda}).$$

The realized number of acquired interactions for this sample is

$$\tilde{T}_i = \sum_{t=1}^{b_i} \xi_i(t).$$

Consequently,

$$\begin{aligned}
\mathbb{E}\Big[\tilde{T}_i \,\Big|\, X_i, T_i, H_i, \widehat{\lambda}\Big] &= \sum_{t=1}^{b_i} \mathbb{P}\Big(\xi_i(t) = 1 \,\Big|\, X_i, T_i, H_i, \widehat{\lambda}\Big) \\
&= \sum_{t=1}^{b_i}\prod_{j=1}^{t} M_j(S_i(j);\widehat{\lambda}) \\
&= \mathcal{B}(M(S_i;\widehat{\lambda})).
\end{aligned} \tag{79}$$

Each deployment trajectory is independent of the CRC sample conditional on the Phase-I information and has the same distribution as the fresh trajectory. Hence,

$$\mathbb{E}\left[\sum_{i\in\mathcal{I}_{\mathrm{cal2}}} \tilde{T}_i\right] = |\mathcal{I}_{\mathrm{cal2}}|\,\mathbb{E}\Big[\mathcal{B}(M(S_\star;\widehat{\lambda}))\Big]. \tag{80}$$

Combining equation 77, equation 78, and equation 80, we obtain

$$\mathbb{E}\left[\sum_{i\in\mathcal{I}_{\mathrm{crc}}} b_i + \sum_{i\in\mathcal{I}_{\mathrm{cal2}}} \tilde{T}_i\right] \le B_{\mathrm{rem}}.$$

Finally, the Phase-I budget consumption is fixed under the conditioning and is equal to $\sum_{i\in\mathcal{I}_{\mathrm{cal1}}} b_i$. Therefore,

$$\begin{aligned}
\mathbb{E}\left[\sum_{i\in\mathcal{I}_{\mathrm{cal}}} \tilde{T}_i\right] &= \sum_{i\in\mathcal{I}_{\mathrm{cal1}}} b_i + \mathbb{E}\left[\sum_{i\in\mathcal{I}_{\mathrm{crc}}} b_i + \sum_{i\in\mathcal{I}_{\mathrm{cal2}}} \tilde{T}_i\right] \\
&\le \sum_{i\in\mathcal{I}_{\mathrm{cal1}}} b_i + B_{\mathrm{rem}} \\
&= B.
\end{aligned}$$

This proves the desired expected budget guarantee conditional on the Phase-I information. Taking expectation over the Phase-I data $\mathcal{D}_{\mathrm{cal1}}$ and $\mathcal{D}_{\mathrm{train}}$ completes the proof. □

C.3 BUDGET VALIDITY OF DAPRO WITHOUT CRC UNDER POLICY GENERALIZATION

In this section, we prove that the expected budget used by DAPRO without CRC, as outlined in Algorithm 3, is bounded by the nominal budget constraint B. Unless stated otherwise, all expectations and probabilities in this section are conditional on $\mathcal{D}_{\text{train}}$; we omit this explicit conditioning for notational convenience.

C.3.1 ASSUMPTIONS

We begin by setting the following assumptions.

Assumption 1 (i.i.d. data)**.** The pairs $\{(S_k)\}_{k=1}^N$ are *independent and identically distributed*, each drawn from a joint distribution ν on $\mathbb{R}^{t_{\max}} \times [0,1]^{t_{\max}}$. We denote by μ the marginal distribution of S_k on $\mathbb{R}^{t_{\max}}$ and by μ_t the marginal distribution of the t-th coordinate $S_k(t)$ on $\mathbb{R}$, for $t \in \{1, \dots, t_{\max}\}$.

Assumption 2 (Split independence)**.** The data is randomly split by π, which is a random variable, taking values in the collection of subsets of $\{1, \dots, N\}$ of size N_1, distributed uniformly over all $\binom{N}{N_1}$ such subsets. The random split π is independent of the full dataset $\{(X_k, T_k, H_k, S_k)\}_{k=1}^N$.

Recall that for a given split π, Algorithm 3 splits the calibration data, indexed by $\{1, \dots, N\}$, as follows:

$$\begin{aligned}
&\mathcal{I}_{\text{cal1}} := \pi \subset \{1, \dots, N\}, \quad |\mathcal{I}_{\text{cal1}}| = N_1, \\
&\mathcal{I}_{\text{cal2}} := \{1, \dots, N\} \setminus \mathcal{I}_{\text{cal1}}, \quad |\mathcal{I}_{\text{cal2}}| = N_2, \\
&\mathcal{D}_{\text{cal1}} := \{(X_i, H_i, T_i)\}_{i \in \mathcal{I}_{\text{cal1}}} \qquad \text{(calibration dataset)}, \\
&S_{\text{cal2}} := S_j \quad \text{for a fixed } j \in \mathcal{I}_{\text{cal2}}.
\end{aligned}$$

By symmetry of the i.i.d. assumption and the uniform split, the choice of j within $\mathcal{I}_{\text{cal2}}$ does not affect any expectation below. We re-index the calibration (cal1) samples as $i \in \{1, \dots, N_1\}$ for notational convenience.

Assumption 3 (Oracle mapping)**.** For each $\beta \geq 0$ and each $t \in \{1, \dots, t_{\max}\}$, there exists a measurable function

$$M_t^*(\,\cdot\,; \beta) : \mathbb{R} \longrightarrow [0,1]$$

that is monotonically increasing in its first argument. For $s = (s(1), \dots, s(t_{\max})) \in \mathbb{R}^{t_{\max}}$ we write $M^*(s; \beta) := (M_1^*(s(1); \beta), \dots, M_{t_{\max}}^*(s(t_{\max}); \beta))$.

Recall that $\bar{B}_2 = (B - B_1)/N_2$ is the remaining budget per sample for Phase II, which is a deterministic function of $\mathcal{D}_{\text{cal1}}$. For a given realization of $\mathcal{D}_{\text{cal1}}$ (and hence a given value of $\bar{B}_2$), the oracle mapping $M^*(\,\cdot\,; \bar{B}_2)$ satisfies the following two conditions $\mathcal{D}_{\text{cal1}}$-almost surely:

1. **(Pathwise oracle fit)** For all $i \in \{1, \dots, N_1\}$:

$$P_i(t) \;=\; M_t^*(S_i(t); \bar{B}_2) \quad \mathbb{P}\text{-almost surely.}$$

2. **(Pathwise population budget constraint)**

$$\mathbb{E}_{S \sim \mu}\big[\mathcal{B}(M^*(S; \bar{B}_2))\big] \;\leq\; \bar{B}_2.$$

Assumption 4 (Estimated mapping)**.** Given $\mathcal{D}_{\text{cal1}}$, the estimated mapping $M_t(\cdot; \mathcal{D}_{\text{cal1}}) : \mathbb{R} \to [0,1]$ is measurable and monotonically increasing for each $t \in \{1, \dots, t_{\max}\}$. Write $M(s; \mathcal{D}_{\text{cal1}}) := (M_1(s(1); \mathcal{D}_{\text{cal1}}), \dots, M_{t_{\max}}(s(t_{\max}); \mathcal{D}_{\text{cal1}}))$.

Assumption 5 (vanishing cross-term moments)**.** For each $t \in \{1, \dots, t_{\max}\}$ and each non-empty subset $A \subseteq \{1, \dots, t\}$, the following holds for $\mathcal{D}_{\text{cal1}}$-almost every realization:

$$\mathbb{E}_{S_{\text{cal2}}}\left[\prod_{k \in A} \varepsilon_k(S_{\text{cal2}}(k); \mathcal{D}_{\text{cal1}}) \cdot \prod_{k \in \{1,\dots,t\} \setminus A} M_k^*(S_{\text{cal2}}(k); \mathcal{D}_{\text{cal1}}) \;\middle|\; \mathcal{D}_{\text{cal1}}\right] \;=\; 0,$$

where $\varepsilon_k(s; \mathcal{D}_{\text{cal1}}) = M_k(s; \mathcal{D}_{\text{cal1}}) - M_k^*(s; \bar{B}_2)$. Observe that for samples i, for times $t > b_i$, $\varepsilon_t(s; \mathcal{D}_{\text{cal1}}) = M_t(s; \mathcal{D}_{\text{cal1}}) = M_t^*(s; \bar{B}_2) = 0$.

Conceptually, Assumption 5 requires that these mixed products have zero conditional expectation over S_{cal2}, given the calibration data $\mathcal{D}_{\text{cal1}}$. We emphasize that this condition is not a mere mathematical artifact of our proof; rather, it is required to guarantee that the projection inaccuracies do not inflate the expected budget consumed during Phase II.

In practice, there are two cases where this assumption is satisfied. First, it holds exactly in finite samples if the coordinates of the score vector S_{cal2} are mutually independent, and the estimated mapping is coordinate-wise unbiased, i.e., $\mathbb{E}_{S_{\text{cal2}}(k)}\left[\varepsilon_k(S_{\text{cal2}}(k); \mathcal{D}_{\text{cal1}})\right] = 0$ almost surely, for all k.

Alternatively, even without assuming independence or finite-sample unbiasedness, this assumption holds asymptotically, assuming $M_t(\cdot \mid \mathcal{D}_{\text{cal1}})$ is a consistent estimator of M_t^*. Specifically, if $\mathbb{E}_{S_{\text{cal2}}(j)}[|M_j(S_{\text{cal2}}(j); \mathcal{D}_{\text{cal1}}) - M_j^*(S_{\text{cal2}}(j))| \mid \mathcal{D}_{\text{cal1}}] \to 0$ as $N_1 \to \infty$, $\mathcal{D}_{\text{cal1}}$-almost surely, then, for any fixed non-empty set A and any $k \in A$, the corresponding cross-term satisfies

$$\left|\mathbb{E}_{S_{\text{cal2}}}\left[\prod_{j\in A}\varepsilon_j \prod_{j\notin A} M_j^*(S_{\text{cal2}}; \bar{B}_2)\middle|\mathcal{D}_{\text{cal1}}\right]\right| \le \mathbb{E}_{S_{\text{cal2}}}\left[\prod_{j\in A}|\varepsilon_j|\middle|\mathcal{D}_{\text{cal1}}\right] \le \mathbb{E}_{S_{\text{cal2}}(k)}\left[|\varepsilon_k||\mathcal{D}_{\text{cal1}}\right] \overset{N_1\to\infty}{\to} 0 \tag{81}$$

where the first inequality uses $M_j^* \in [0,1]$, and the second uses $|\varepsilon_j| \le 1$ for all j. Thus, Assumption 5 holds asymptotically for any fixed $\mathcal{D}_{\text{cal1}}$ and target budget $\bar{B}_2$, meaning the budget constraint can be satisfied with arbitrary precision given sufficient calibration data in Phase I.

C.3.2 Expected budget validity proofs

To establish the expected budget bound, we first prove the following lemmas.

Independence lemma

Lemma 1 (Independence of $\mathcal{D}_{\text{cal1}}$ and S_{cal2})**.** *Under Assumptions 1 and 2,*

$$\mathcal{D}_{\text{cal1}} \;\perp\; S_{\text{cal2}}.$$

Proof. Given the split π, $\mathcal{D}_{\text{cal1}}$ and S_{cal2} correspond to disjoint subsets of the i.i.d. array $\{(S_k, P_k)\}_{k=1}^N$. Disjoint subsets of an i.i.d. sequence are independent of each other. Since π is independent of the data (Assumption 2), conditioning on π does not introduce dependence, and so $\mathcal{D}_{\text{cal1}} \perp S_{\text{cal2}}$. □

Properties of the budget function Throughout this section we adopt the convention that the continuation probability of a trajectory vanishes beyond its acquisition horizon:

$$P_i(t) \;=\; M_t(S_i(t))\,\mathbb{I}\{t \le b_i\}, \qquad P_i^*(t) \;=\; M_t^*(S_i(t); \bar{B}_2)\,\mathbb{I}\{t \le b_i\}. \tag{82}$$

The budget function $\mathcal{B} : [0,1]^{t_{\max}} \to \mathbb{R}$ is given by:

$$\mathcal{B}(P) \;=\; \sum_{t=1}^{\min(T_i, \hat{f}_{\text{prior}}(X_i))} \prod_{j=1}^{t} P(j), \qquad P = (P(1), \ldots, P(t_{\max})) \in [0,1]^{t_{\max}}.$$

Lemma 2 (Properties of the budget function)**.**

1. ***Boundedness:*** $0 \le \mathcal{B}(P) \le t_{\max}$ *for all* $P \in [0,1]^{t_{\max}}$.
2. ***Multilinearity:*** $\mathcal{B}$ *is a polynomial in* $\{P(j)\}_{j=1}^{t_{\max}}$ *that is linear in each coordinate* $P(j)$ *when all other coordinates are held fixed.*
3. ***Monotonicity:*** $\mathcal{B}$ *is monotonically increasing in each coordinate* $P(j)$.

Proof. (i) Each product $\prod_{j=1}^t P(j) \in [0,1]$ and there are $t_{\max}$ terms. (ii) Direct inspection of the polynomial structure. (iii) For any j_0, increasing $P(j_0)$ increases every term $\prod_{j=1}^t P(j)$ with $t \ge j_0$. □

Estimation error does not affect expected loss

Lemma 3 (Cross-term cancellation)**.** *Under Assumptions 1, 2, 3, 4, and 5, the following holds for $\mathcal{D}_{\mathrm{cal1}}$-almost every realization of the calibration data:*

$$\mathbb{E}_{S_{\mathrm{cal2}}}\big[\mathcal{B}(M(S_{\mathrm{cal2}};\mathcal{D}_{\mathrm{cal1}}))\mid\mathcal{D}_{\mathrm{cal1}}\big] \;=\; \mathbb{E}_{S_{\mathrm{cal2}}}\big[\mathcal{B}(M^*(S_{\mathrm{cal2}};\bar{B}_2))\mid\mathcal{D}_{\mathrm{cal1}}\big].$$

Here $\mathbb{E}_{S_{\mathrm{cal2}}}[\,\cdot\mid\mathcal{D}_{\mathrm{cal1}}]$ denotes the conditional expectation over S_{cal2} given $\mathcal{D}_{\mathrm{cal1}}$. Since $\mathcal{D}_{\mathrm{cal1}}\perp S_{\mathrm{cal2}}$ (Lemma 1), this conditional expectation equals the marginal expectation over $S_{\mathrm{cal2}}\sim\mu$, with $\mathcal{D}_{\mathrm{cal1}}$ treated as a fixed parameter.

Proof. Fix a realization of $\mathcal{D}_{\mathrm{cal1}}$ (a.s.). Define for each $j\in\{1,\dots,t_{\max}\}$:

$$\begin{aligned} p_j := M_j^*(S_{\mathrm{cal2}}(j);\bar{B}_2), \qquad e_j &:= \varepsilon_j(S_{\mathrm{cal2}}(j);\mathcal{D}_{\mathrm{cal1}}) = M_j(S_{\mathrm{cal2}}(j);\mathcal{D}_{\mathrm{cal1}}) \\ &\quad - M_j^*(S_{\mathrm{cal2}}(j);\bar{B}_2). \end{aligned} \tag{83}$$

Given $\mathcal{D}_{\mathrm{cal1}}$, both p_j and e_j are bounded, measurable functions of $S_{\mathrm{cal2}}(j)$ alone. For each $t\in\{1,\dots,t_{\max}\}$, expand each factor as $M_j(S_{\mathrm{cal2}}(j);\mathcal{D}_{\mathrm{cal1}}) = p_j+e_j$ and expand the product by the distributive law:

$$\prod_{j=1}^{t} M_j(S_{\mathrm{cal2}}(j);\mathcal{D}_{\mathrm{cal1}}) = \prod_{j=1}^{t}(p_j+e_j) = \sum_{A\subseteq\{1,\dots,t\}}\left(\prod_{j\in A} e_j\right)\left(\prod_{j\in\{1,\dots,t\}\setminus A} p_j\right). \tag{84}$$

We prove the above by induction over t. For $t=1$: $\prod_{j=1}^{1}(p_j+e_j) = p_1+e_1$, which satisfies this identity. Assume that the identity holds for some arbitrary integer $k\geq 1$. Then, for $t=k+1$:

$$\prod_{j=1}^{k+1}(p_j+e_j) = \left(\prod_{j=1}^{k}(p_j+e_j)\right)(p_{k+1}+e_{k+1})$$

By our inductive assumption, we can substitute the expansion for the product up to k:

$$= \left(\sum_{A\subseteq S_k}\left(\prod_{j\in A} e_j\right)\left(\prod_{j\in S_k\setminus A} p_j\right)\right)(p_{k+1}+e_{k+1})$$

Distributing the terms p_{k+1} and e_{k+1} across the sum gives two separate sums:

$$\begin{aligned} &= \sum_{A\subseteq S_k}\left(\prod_{j\in A} e_j\right)\left(\prod_{j\in S_k\setminus A} p_j\right)p_{k+1} \;+\; \sum_{A\subseteq S_k}\left(\prod_{j\in A} e_j\right)\left(\prod_{j\in S_k\setminus A} p_j\right)e_{k+1} \\ &= \sum_{X\subseteq S_{k+1}}\left(\prod_{j\in X} e_j\right)\left(\prod_{j\in S_{k+1}\setminus X} p_j\right) \\ &= \prod_{j=1}^{k+1}(p_j+e_j). \end{aligned}$$

We now take the conditional expectation given $\mathcal{D}_{\mathrm{cal1}}$.

$$\begin{aligned} &\mathbb{E}_{S_{\mathrm{cal2}}}\left[\prod_{j=1}^{t} M_j(S_{\mathrm{cal2}}(j);\mathcal{D}_{\mathrm{cal1}}) \;\middle|\; \mathcal{D}_{\mathrm{cal1}}\right] \\ &= \sum_{A\subseteq\{1,\dots,t\}} \mathbb{E}_{S_{\mathrm{cal2}}}\left[\prod_{j\in A} e_j\cdot\prod_{j\in\{1,\dots,t\}\setminus A} p_j \;\middle|\; \mathcal{D}_{\mathrm{cal1}}\right]. \end{aligned} \tag{85}$$

Exchanging the sum and expectation is valid since the sum is finite and all terms are bounded (Lemma 2(i)). Next, we isolate the $A=\emptyset$ term.

$$equation\ 85 = \underbrace{\mathbb{E}_{S_{\text{cal2}}}\left[\prod_{j=1}^{t} p_j \;\middle|\; \mathcal{D}_{\text{cal1}}\right]}_{A=\emptyset} + \sum_{\emptyset \neq A \subseteq \{1,\ldots,t\}} \mathbb{E}_{S_{\text{cal2}}}\left[\prod_{j\in A} e_j \cdot \prod_{j\in\{1,\ldots,t\}\setminus A} p_j \;\middle|\; \mathcal{D}_{\text{cal1}}\right]. \tag{86}$$

Now, for each non-empty $A \subseteq \{1,\ldots,t\}$, we apply Assumption 5:

$$\mathbb{E}_{S_{\text{cal2}}}\left[\prod_{j\in A} e_j \cdot \prod_{j\in\{1,\ldots,t\}\setminus A} p_j \;\middle|\; \mathcal{D}_{\text{cal1}}\right] = 0 \qquad \mathcal{D}_{\text{cal1}}\text{-a.s.}$$

Hence the entire sum over non-empty A in equation 86 is zero, and:

$$\mathbb{E}_{S_{\text{cal2}}}\left[\prod_{j=1}^{t} M_j(S_{\text{cal2}}(j); \mathcal{D}_{\text{cal1}}) \;\middle|\; \mathcal{D}_{\text{cal1}}\right] = \mathbb{E}_{S_{\text{cal2}}}\left[\prod_{j=1}^{t} M_j^*(S_{\text{cal2}}(j); \bar{B}_2) \;\middle|\; \mathcal{D}_{\text{cal1}}\right]. \tag{87}$$

Summing equation 87 over $t \in \{1,\ldots,t_{\max}\}$ ($P_i(t) = M_t^*(S_i; \bar{B}_2) = 0$ for $t > b_i$) and using linearity of conditional expectation:

$$\mathbb{E}_{S_{\text{cal2}}}\big[\mathcal{B}(M(S_{\text{cal2}}; \mathcal{D}_{\text{cal1}})) \mid \mathcal{D}_{\text{cal1}}\big] = \mathbb{E}_{S_{\text{cal2}}}\big[\mathcal{B}(M^*(S_{\text{cal2}}; \bar{B}_2)) \mid \mathcal{D}_{\text{cal1}}\big].$$

□

Oracle budget is bounded

Lemma 4 (Oracle budget bound)**.** *Under Assumption 3, for $\mathcal{D}_{\text{cal1}}$-almost every realization:*

$$\mathbb{E}_{S_{\text{cal2}}}\big[\mathcal{B}(M^*(S_{\text{cal2}}; \bar{B}_2)) \mid \mathcal{D}_{\text{cal1}}\big] \;\le\; \bar{B}_2.$$

Proof. This follows directly from condition (ii) of Assumption 3, which states that $\mathbb{E}_{S_{\text{cal2}}}[\mathcal{B}(M^*(S_{\text{cal2}}; \bar{B}_2)) \mid \mathcal{D}_{\text{cal1}}] \le \bar{B}_2$ holds $\mathcal{D}_{\text{cal1}}$-almost surely. □

Proof of the Holdout Budget Bound

Proposition 3 (Holdout sample budget bound)**.** *Under Assumptions 1, 2, 3, 4, and 5, for $\mathcal{D}_{\text{cal1}}$-almost every realization:*

$$\mathbb{E}_{S_{\text{cal2}}}\big[\mathcal{B}(M(S_{\text{cal2}}; \mathcal{D}_{\text{cal1}})) \mid \mathcal{D}_{\text{cal1}}\big] \;\le\; \bar{B}_2.$$

Proof of Proposition 3. By Lemma 3 (cross-term cancellation), for $\mathcal{D}_{\text{cal1}}$-almost every realization:

$$\mathbb{E}_{S_{\text{cal2}}}\big[\mathcal{B}(M(S_{\text{cal2}}; \mathcal{D}_{\text{cal1}})) \mid \mathcal{D}_{\text{cal1}}\big] \;=\; \mathbb{E}_{S_{\text{cal2}}}\big[\mathcal{B}(M^*(S_{\text{cal2}}; \bar{B}_2)) \mid \mathcal{D}_{\text{cal1}}\big].$$

By Lemma 4 (oracle budget bound), for $\mathcal{D}_{\text{cal1}}$-almost every realization:

$$\mathbb{E}_{S_{\text{cal2}}}\big[\mathcal{B}(M^*(S_{\text{cal2}}; \bar{B}_2)) \mid \mathcal{D}_{\text{cal1}}\big] \;\le\; \bar{B}_2.$$

Chaining these two relations gives the result. □

Proof of the Algorithm's Budget Validity

Theorem 6 (Algorithm's budget bound)**.** *Under Assumptions 1, 2, 3, 4, and 5, the expected budget used by Algorithm 3 is bounded by B.*

Proof of Theorem 6. Condition on $\mathcal{D}_{\text{cal1}}$. Given $\mathcal{D}_{\text{cal1}}$, the Phase I budget $B_1 = \sum_{i\in\mathcal{I}_{\text{cal1}}} b_i$ is a fixed constant, and $\bar{B}_2 = (B - B_1)/N_2$ is a fixed constant as well. The total realized budget is $B_1 + B_2$, where $B_2 = \sum_{i\in\mathcal{I}_{\text{cal2}}} U_i$ is random over Phase II draws, with $U_i = \sum_{t=1}^{b_i} \prod_{j=1}^{t} \xi_i(j)$.

By linearity of conditional expectation:

$$\mathbb{E}[B_2 \mid \mathcal{D}_{\text{cal1}}] = \sum_{i\in\mathcal{I}_{\text{cal2}}} \mathbb{E}[U_i \mid \mathcal{D}_{\text{cal1}}]. \tag{88}$$

For each $i \in \mathcal{I}_{\text{cal2}}$, by the tower property and $\mathcal{D}_{\text{cal1}} \perp\!\!\!\perp S_i$ (Lemma 1):

$$\mathbb{E}[U_i \mid \mathcal{D}_{\text{cal1}}] = \mathbb{E}_{S_i}\Big[\mathbb{E}[U_i \mid S_i, T_i, \mathcal{D}_{\text{cal1}}]\Big|\mathcal{D}_{\text{cal1}}\Big] = \mathbb{E}_{S_{\text{cal2}}}[\mathcal{B}(M(S_{\text{cal2}}; \mathcal{D}_{\text{cal1}}))]\,, \tag{89}$$

where the second equality uses $\mathbb{E}[U_i \mid S_i, T_i, \mathcal{D}_{\text{cal1}}] = \mathcal{B}(M(S_i; \mathcal{D}_{\text{cal1}}))$ (the expected stopping time equals the budget function evaluated at the continuation probabilities, given the Bernoulli draws are independent).

By Proposition 3, for $\mathcal{D}_{\text{cal1}}$-almost every realization:

$$\mathbb{E}_{S_{\text{cal2}}}[\mathcal{B}(M(S_{\text{cal2}}; \mathcal{D}_{\text{cal1}})) \mid \mathcal{D}_{\text{cal1}}] \le \bar{B}_2. \tag{90}$$

Therefore $\mathbb{E}[B_2 \mid \mathcal{D}_{\text{cal1}}] \le N_2 \bar{B}_2$, and:

$$\mathbb{E}[B_1 + B_2 \mid \mathcal{D}_{\text{cal1}}] \le B_1 + N_2 \bar{B}_2 \tag{91}$$

$$= B_1 + N_2 \cdot \frac{B - B_1}{N_2} \tag{92}$$

$$= B. \tag{93}$$

Taking the expectation over $\mathcal{D}_{\text{cal1}}$ gives $\mathbb{E}[B_1 + B_2] \le B$. □

C.3.3 Finite-sample budget guarantee

In this section, we provide a finite-sample bound for the budget consumed by Algorithm 3. Throughout this section, all probabilities and expectations are conditioned on $\mathcal{D}_{\text{train}}$. For notational convenience, we omit this conditioning.

Theorem 7. *Under Assumptions 1, 2, 4, 3, and 5, with probability at least $1 - \delta$, the average budget per sample consumed by Algorithm 3 is bounded by:*

$$\frac{1}{N}\sum_{i \in \mathcal{I}_{\text{cal}}} \tilde{T}_i \le \frac{B}{N} + \frac{t_{\max}\log(1/\delta)}{3N} + \frac{1}{N}\sqrt{\frac{{t_{\max}}^2 \log^2(1/\delta)}{9} + 2N_2 t_{\max}\bar{B}_2 \log(1/\delta)}. \tag{94}$$

Proof. We denote $B_1 = \sum_{i \in \mathcal{I}_{\text{cal1}}} b_i$, where $b_i = \min(T_i, \hat{f}_{\text{prior}}(X_i))$, and $U_i = \sum_{t=1}^{b_i} \prod_{j=1}^{t} Z_j^i$ for any $i \in \mathcal{I}_{\text{cal2}}$, so that $B_2 = \sum_{i \in \mathcal{I}_{\text{cal2}}} U_i$.

For a fixed $\mathcal{D}_{\text{cal1}}$, and fixed (S_i, T_i), the continuation probabilities $M_j(S_i(j); \mathcal{D}_{\text{cal1}})$ are deterministic. The Bernoulli draws $\{Z_t^i\}_t$ are then mutually independent, so:

$$\mathbb{E}\left[U_i \mid \mathcal{D}_{\text{cal1}}, S_i, T_i\right] = \sum_{t=1}^{b_i}\prod_{j=1}^{t} M_j(S_i(j); \mathcal{D}_{\text{cal1}}) = \mathcal{B}(M(S_i; \mathcal{D}_{\text{cal1}})). \tag{95}$$

Taking a further expectation over (S_i, T_i) and using $\mathcal{D}_{\text{cal1}} \perp\!\!\!\perp S_i$ (Lemma 1):

$$\mathbb{E}\left[U_i \mid \mathcal{D}_{\text{cal1}}\right] = \mathbb{E}_{S_{\text{cal2}}}[\mathcal{B}(M(S_{\text{cal2}}; \mathcal{D}_{\text{cal1}})) \mid \mathcal{D}_{\text{cal1}}]\,. \tag{96}$$

By Proposition 3, for $\mathcal{D}_{\text{cal1}}$-almost every realization:

$$\mathbb{E}\left[U_i \mid \mathcal{D}_{\text{cal1}}\right] = \mathbb{E}_{S_{\text{cal2}}}[\mathcal{B}(M(S_{\text{cal2}}; \mathcal{D}_{\text{cal1}})) \mid \mathcal{D}_{\text{cal1}}] \le \bar{B}_2. \tag{97}$$

Hence:

$$\mathbb{E}\left[B_2 \mid \mathcal{D}_{\text{cal1}}\right] = \sum_{i \in \mathcal{I}_{\text{cal2}}} \mathbb{E}\left[U_i \mid \mathcal{D}_{\text{cal1}}\right] \le N_2 \bar{B}_2. \tag{98}$$

Notice that the collection $\{U_i\}_{i \in \mathcal{I}_{\text{cal2}}}$ is mutually independent given $\mathcal{D}_{\text{cal1}}$. Since $U_i \in [0, t_{\max}]$ almost surely, by Lemma 2, and $\mathbb{E}[U_i \mid \mathcal{D}_{\text{cal1}}] \le \bar{B}_2$, the Bhatia–Davis inequality gives:

$$\text{Var}[U_i \mid \mathcal{D}_{\text{cal1}}] \le (t_{\max} - \mathbb{E}[U_i \mid \mathcal{D}_{\text{cal1}}]) \cdot \mathbb{E}[U_i \mid \mathcal{D}_{\text{cal1}}] \le t_{\max}\bar{B}_2. \tag{99}$$

Observe that:

$$\mathbb{E}\left[\left(U_i - \frac{1}{N_2}\mathbb{E}[B_2 \mid \mathcal{D}_{\text{cal1}}]\right)^2 \middle| \mathcal{D}_{\text{cal1}}\right] = \mathbb{E}[(U_i - \mathbb{E}[U_i \mid \mathcal{D}_{\text{cal1}}])^2 \mid \mathcal{D}_{\text{cal1}}] = \text{Var}(U_i \mid \mathcal{D}_{\text{cal1}}). \tag{100}$$

We now apply the Bernstein inequality to the sum $B_2 = \sum_{i \in \mathcal{I}_{\text{cal2}}} U_i$. Notice that $\{U_i\}_{i \in \mathcal{I}_{\text{cal2}}}$ are mutually independent variables given $\mathcal{D}_{\text{cal1}}$, where each $U_i \in [0, t_{\max}]$.

$$\begin{aligned}\mathbb{P}(B_2 - \mathbb{E}[B_2 \mid \mathcal{D}_{\text{cal1}}] \geq \varepsilon \mid \mathcal{D}_{\text{cal1}}) &\leq \exp\left(-\frac{\varepsilon^2/2}{\sum_{i \in \mathcal{I}_{\text{cal2}}} \text{Var}(U_i \mid \mathcal{D}_{\text{cal1}}) + t_{\max}\varepsilon/3}\right) \\ &\leq \exp\left(-\frac{\varepsilon^2/2}{N_2 t_{\max} \bar{B}_2 + t_{\max}\varepsilon/3}\right).\end{aligned} \tag{101}$$

We now show that by setting the ε as given below, the probability above is bounded by δ:

$$\varepsilon = \frac{t_{\max}\log(1/\delta)}{3} + \sqrt{\frac{{t_{\max}}^2 \log^2(1/\delta)}{9} + 2N_2 t_{\max}\bar{B}_2 \log(1/\delta)} \tag{102}$$

For ease of notation, we denote $V = \sum_{i \in \mathcal{I}_{\text{cal2}}} t_{\max}\bar{B}_2$.

$$\varepsilon - \frac{t_{\max}\log(1/\delta)}{3} = \sqrt{\frac{{t_{\max}}^2 \log^2(1/\delta)}{9} + 2V\log(1/\delta)} \tag{103}$$

Squaring both sides yields:

$$\left(\varepsilon - \frac{t_{\max}\log(1/\delta)}{3}\right)^2 = \frac{{t_{\max}}^2 \log^2(1/\delta)}{9} + 2V\log(1/\delta) \tag{104}$$

$$\varepsilon^2 - \frac{2t_{\max}\log(1/\delta)}{3}\varepsilon + \frac{{t_{\max}}^2 \log^2(1/\delta)}{9} = \frac{{t_{\max}}^2 \log^2(1/\delta)}{9} + 2V\log(1/\delta) \tag{105}$$

Notice that the $\frac{{t_{\max}}^2 \log^2(1/\delta)}{9}$ term appears on both sides and cancels out:

$$\varepsilon^2 - \frac{2t_{\max}\log(1/\delta)}{3}\varepsilon = 2V\log(1/\delta) \tag{106}$$

Divide the entire equation by 2:

$$\frac{\varepsilon^2}{2} - \frac{t_{\max}\varepsilon}{3}\log(1/\delta) = V\log(1/\delta) \tag{107}$$

Move the ε term to the right side:

$$\frac{\varepsilon^2}{2} = V\log(1/\delta) + \frac{t_{\max}\varepsilon}{3}\log(1/\delta) \tag{108}$$

$$\frac{\varepsilon^2}{2} = \left(V + \frac{t_{\max}\varepsilon}{3}\right)\log(1/\delta) \tag{109}$$

The exponent in the original inequality bound is:

$$-\frac{\varepsilon^2/2}{V + t_{\max}\varepsilon/3}. \tag{110}$$

Substituting the equivalent expression we just found for $\frac{\varepsilon^2}{2}$:

$$-\frac{\left(V + \frac{t_{\max}\varepsilon}{3}\right)\log(1/\delta)}{V + t_{\max}\varepsilon/3} = -\log(1/\delta). \tag{111}$$

Therefore,

$$\exp(-\log(1/\delta)) = \exp(\log(\delta)) = \delta. \tag{112}$$

Putting it all together into a single block to match the original formulation:

$$\begin{aligned}\mathbb{P}(B_2 - \mathbb{E}[B_2 \mid \mathcal{D}_{\text{cal1}}] \geq \varepsilon \mid \mathcal{D}_{\text{cal1}}) &\leq \exp\left(-\frac{\varepsilon^2/2}{\sum_{i\in\mathcal{I}_{\text{cal2}}} t_{\max}\bar{B}_2 + t_{\max}\varepsilon/3}\right) \\ &= \exp\left(-\frac{\left(\sum_{i\in\mathcal{I}_{\text{cal2}}} t_{\max}\bar{B}_2 + \frac{t_{\max}\varepsilon}{3}\right)\log(1/\delta)}{\sum_{i\in\mathcal{I}_{\text{cal2}}} t_{\max}\bar{B}_2 + t_{\max}\varepsilon/3}\right) \\ &= \exp\left(-\log(1/\delta)\right) \\ &= \delta.\end{aligned} \tag{113}$$

Therefore, with probability at least $1-\delta$, the budget used in Phase II is bounded by:

$$B_2 \leq N_2\bar{B}_2 + \frac{t_{\max}\log(1/\delta)}{3} + \sqrt{\frac{{t_{\max}}^2\log^2(1/\delta)}{9} + 2N_2 t_{\max}\bar{B}_2\log(1/\delta)}. \tag{114}$$

By the construction of $\bar{B}_2$ of Algorithm 3, we have:

$$B_1 + N_2\bar{B}_2 \leq B. \tag{115}$$

Therefore, the average budget per sample used by Algorithm 3 is bounded by:

$$\frac{1}{N}\sum_{i\in\mathcal{I}_{\text{cal}}}\tilde{T}_i \leq \frac{B}{N} + \frac{t_{\max}\log(1/\delta)}{3N} + \frac{1}{N}\sqrt{\frac{{t_{\max}}^2\log^2(1/\delta)}{9} + 2N_2 t_{\max}\bar{B}_2\log(1/\delta)}. \tag{116}$$

□

C.3.4 Budget Guarantee with Bounded Estimation Errors

We derive an approximate budget guarantee under an alternative condition that bounds the mean absolute projection error at each step, without requiring exact cancellation of the mixed terms. Specifically, we rely on the following assumption:

Assumption 6 (Bounded Mean Absolute Estimation Error)**.** For each $j \in \{1, \ldots, t_{\max}\}$, there exists a constant $\eta_j \geq 0$ such that, almost surely over $\mathcal{D}_{\text{cal1}}$:

$$\mathbb{E}_{S_{\text{cal2}}(j)}[|\varepsilon_j(S_{\text{cal2}}(j); \mathcal{D}_{\text{cal1}})| \mid \mathcal{D}_{\text{cal1}}] \leq \eta_j, \tag{117}$$

where $\varepsilon_j(s; \mathcal{D}_{\text{cal1}}) := M_j(s; \mathcal{D}_{\text{cal1}}) - M_j^*(s; \bar{B}_2)$ is the pointwise estimation error at step j.

Observe that Assumption 5 requires the cross-term expectations:

$$\mathbb{E}[\prod_{j\in A}\varepsilon_j(s; \mathcal{D}_{\text{cal1}}) \cdot \prod_{j\notin A} M_j^*(s; \bar{B}_2) \mid \mathcal{D}_{\text{cal1}}] = 0$$

for every non-empty A, which implies $\eta_j = 0$ in case where S_{cal2} has independent coordinates. In contrast, Assumption 6 is different: it requires only that the marginal mean absolute error at each coordinate is bounded, without making assumptions on the joint cross-term. In particular, it does not require that the score coordinates are independent. Furthermore, this assumption is reasonable for consistent estimators $M_j(\cdot; \mathcal{D}_{\text{cal1}})$ that satisfy $\mathbb{E}_{S_{\text{cal2}}(j)}[|\varepsilon_j(S_{\text{cal2}}(j); \mathcal{D}_{\text{cal1}})| \mid \mathcal{D}_{\text{cal1}}] \to 0$ as $N_1 \to \infty$, and thus $\eta_j \to 0$.

Theorem 8 (Expected Budget with Estimation Error)**.** *Under Assumptions 1, 2, 3, 4 and 6, the expected total budget consumed by Algorithm 3 satisfies:*

$$\mathbb{E}_{\mathcal{D}_{\text{cal}}}\left[\sum_{i\in\mathcal{I}_{\text{cal}}}\tilde{T}_i\right] \;\leq\; B + N_2\,\Gamma, \tag{118}$$

where $\Gamma = \sum_{k=1}^{t_{\max}}(t_{\max} - k + 1)\eta_k$.

To prove it, we first develop an upper bound on the expected budget used in Phase II:

Lemma 5. *Under Assumptions 1, 2, 3, 4, for $\mathcal{D}_{\text{cal1}}$-almost every realization:*

$$\mathbb{E}_{S_{\text{cal2}}}[\mathcal{B}(M(S_{\text{cal2}};\mathcal{D}_{\text{cal1}})) \mid \mathcal{D}_{\text{cal1}}] \;\le\; \mathbb{E}_{S_{\text{cal2}}}\big[\mathcal{B}(M^*(S_{\text{cal2}};\bar{B}_2))\big] \;+\; \Gamma, \tag{119}$$

where the total bias is:

$$\Gamma \;:=\; \sum_{k=1}^{t_{\max}} (t_{\max}-k+1)\,\eta_k. \tag{120}$$

Proof. We first show a telescoping product identity. For any two sequences of real numbers $\{a_j\}_{j=1}^t$ and $\{b_j\}_{j=1}^t$, the following identity holds:

$$\prod_{j=1}^{t} a_j - \prod_{j=1}^{t} b_j = \sum_{k=1}^{t} \left(\prod_{j=1}^{k-1} a_j\right)(a_k - b_k)\left(\prod_{j=k+1}^{t} b_j\right), \tag{121}$$

where empty products equal 1. We verify by induction. For $t=1$: $a_1 - b_1 = (a_1 - b_1)$, which is trivially true. Assuming it holds for $t-1$:

$$\begin{aligned}
\prod_{j=1}^{t} a_j - \prod_{j=1}^{t} b_j &= a_t \prod_{j=1}^{t-1} a_j - b_t \prod_{j=1}^{t-1} b_j \\
&= (a_t - b_t)\prod_{j=1}^{t-1} a_j + b_t\left(\prod_{j=1}^{t-1} a_j - \prod_{j=1}^{t-1} b_j\right) \\
&= (a_t - b_t)\prod_{j=1}^{t-1} a_j + b_t \sum_{k=1}^{t-1}\left(\prod_{j=1}^{k-1} a_j\right)(a_k - b_k)\left(\prod_{j=k+1}^{t-1} b_j\right) \\
&= \sum_{k=1}^{t}\left(\prod_{j=1}^{k-1} a_j\right)(a_k - b_k)\left(\prod_{j=k+1}^{t} b_j\right),
\end{aligned} \tag{122}$$

completing the induction.

We now apply this identity to the continuation probabilities. Set $a_j = M_j(S_{\text{cal2}}(j);\mathcal{D}_{\text{cal1}})$ and $b_j = M_j^*(S_{\text{cal2}}(j);\bar{B}_2)$, so $a_j - b_j = \varepsilon_j(S_{\text{cal2}}(j);\mathcal{D}_{\text{cal1}})$ and $a_j, b_j \in [0,1]$. Then for each $t \in \{1,\dots,t_{\max}\}$:

$$\begin{aligned}
&\prod_{j=1}^{t} M_j(S_{\text{cal2}}(j);\mathcal{D}_{\text{cal1}}) - \prod_{j=1}^{t} M_j^*(S_{\text{cal2}}(j);\bar{B}_2) \\
&= \sum_{k=1}^{t} \underbrace{\left(\prod_{j=1}^{k-1} M_j(S_{\text{cal2}}(j);\mathcal{D}_{\text{cal1}})\right)}_{=:\,W_k^- \,\in\, [0,1]} \cdot \varepsilon_k(S_{\text{cal2}}(k);\mathcal{D}_{\text{cal1}}) \cdot \underbrace{\left(\prod_{j=k+1}^{t} M_j^*(S_{\text{cal2}}(j);\bar{B}_2)\right)}_{=:\,W_k^+ \,\in\, [0,1]}.
\end{aligned} \tag{123}$$

Since $W_k^-, W_k^+ \in [0,1]$ (being products of values in $[0,1]$), we have $|W_k^- \cdot \varepsilon_k \cdot W_k^+| \le |\varepsilon_k|$ pointwise. Therefore, taking the conditional expectation given $\mathcal{D}_{\text{cal1}}$:

$$\begin{aligned}
\left|\mathbb{E}_{S_{\text{cal2}}}\left[\prod_{j=1}^{t} M_j - \prod_{j=1}^{t} M_j^* \;\middle|\; \mathcal{D}_{\text{cal1}}\right]\right| &\le \mathbb{E}_{S_{\text{cal2}}}\left[\left|\prod_{j=1}^{t} M_j - \prod_{j=1}^{t} M_j^*\right| \;\middle|\; \mathcal{D}_{\text{cal1}}\right] \\
&\le \mathbb{E}_{S_{\text{cal2}}}\left[\sum_{k=1}^{t} W_k^- \,|\varepsilon_k(S_{\text{cal2}}(k);\mathcal{D}_{\text{cal1}})|\, W_k^+ \;\middle|\; \mathcal{D}_{\text{cal1}}\right] \\
&\le \sum_{k=1}^{t} \mathbb{E}_{S_{\text{cal2}}}[|\varepsilon_k(S_{\text{cal2}}(k);\mathcal{D}_{\text{cal1}})| \mid \mathcal{D}_{\text{cal1}}],
\end{aligned} \tag{124}$$

where the final inequality uses $W_k^- W_k^+ \leq 1$ and the fact that $|\varepsilon_k(S_{\text{cal2}}(k); \mathcal{D}_{\text{cal1}})|$ depends only on the k-th coordinate $S(k)$, so that $\mathbb{E}_{S_{\text{cal2}}}[W_k^- |\varepsilon_k(S_{\text{cal2}}(k); \mathcal{D}_{\text{cal1}})| W_k^+ \mid \mathcal{D}_{\text{cal1}}] \leq \mathbb{E}_{S_{\text{cal2}}}[|\varepsilon_k(S_{\text{cal2}}(k); \mathcal{D}_{\text{cal1}})| \mid \mathcal{D}_{\text{cal1}}] = \mathbb{E}_{S_{\text{cal2}}(k)}[|\varepsilon_k(S_{\text{cal2}}(k); \mathcal{D}_{\text{cal1}})| \mid \mathcal{D}_{\text{cal1}}]$ regardless of the joint distribution of coordinates.

Applying Assumption 6: $\mathbb{E}_{S_{\text{cal2}}(k)}[|\varepsilon_k(S_{\text{cal2}}(k); \mathcal{D}_{\text{cal1}})| \mid \mathcal{D}_{\text{cal1}}] \leq \eta_k$, giving:

$$\left| \mathbb{E}_{S_{\text{cal2}}} \left[\prod_{j=1}^{t} M_j(S_{\text{cal2}}; \mathcal{D}_{\text{cal1}}) - \prod_{j=1}^{t} M_j^*(S_{\text{cal2}}; \bar{B}_2) \,\middle|\, \mathcal{D}_{\text{cal1}} \right] \right| \leq \sum_{k=1}^{t} \eta_k. \tag{125}$$

We now sum over t:

$$\begin{aligned}
&\mathbb{E}_{S_{\text{cal2}}}[\mathcal{B}(M(S_{\text{cal2}}; \mathcal{D}_{\text{cal1}})) \mid \mathcal{D}_{\text{cal1}}] - \mathbb{E}_{S_{\text{cal2}}}[\mathcal{B}(M^*(S_{\text{cal2}}; \bar{B}_2)) \mid \mathcal{D}_{\text{cal1}}] \qquad (126)\\
&= \sum_{t=1}^{t_{\max}} \mathbb{E}_{S_{\text{cal2}}} \left[\prod_{j=1}^{t} M_j(S_{\text{cal2}}(j); \mathcal{D}_{\text{cal1}}) - \prod_{j=1}^{t} M_j^*(S_{\text{cal2}}(j); \bar{B}_2) \,\middle|\, \mathcal{D}_{\text{cal1}} \right] \\
&\leq \sum_{t=1}^{t_{\max}} \sum_{k=1}^{t} \eta_k \\
&= \sum_{k=1}^{t_{\max}} \eta_k \cdot |\{t : t \geq k,\, t \leq t_{\max}\}| \\
&= \sum_{k=1}^{t_{\max}} \eta_k (t_{\max} - k + 1) = \Gamma, \qquad (127)
\end{aligned}$$

where we interchanged the order of summation: $\sum_{t=1}^{t_{\max}} \sum_{k=1}^{t} = \sum_{k=1}^{t_{\max}} \sum_{t=k}^{t_{\max}}$, and $\sum_{t=k}^{t_{\max}} 1 = t_{\max} - k + 1$. □

Armed with the above Lemma, we turn to prove Theorem 8.

Proof of Theorem 8. The Phase I budget $B_1 = \sum_{i \in \mathcal{I}_{\text{cal1}}} b_i$ satisfies $\mathbb{E}[B_1] = N_1 \mathbb{E}[b_1]$, and by construction $\bar{B}_2 = (B - B_1)/N_2$ so $B_1 + N_2 \bar{B}_2 = B$ deterministically. For the Phase II budget:

$$\mathbb{E}[B_2 \mid \mathcal{D}_{\text{cal1}}] = N_2 \, \mathbb{E}_{S_{\text{cal2}}}[\mathcal{B}(M(S_{\text{cal2}}; \mathcal{D}_{\text{cal1}})) \mid \mathcal{D}_{\text{cal1}}]. \tag{128}$$

By Lemma 5 and Lemma 4:

$$\mathbb{E}_{S_{\text{cal2}}}[\mathcal{B}(M(S_{\text{cal2}}; \mathcal{D}_{\text{cal1}})) \mid \mathcal{D}_{\text{cal1}}] \leq \mathbb{E}_{S_{\text{cal2}}}[\mathcal{B}(M^*(S_{\text{cal2}}; \bar{B}_2))] + \Gamma \leq \bar{B}_2 + \Gamma. \tag{129}$$

Therefore:

$$\mathbb{E}[B_2 \mid \mathcal{D}_{\text{cal1}}] \leq N_2(\bar{B}_2 + \Gamma) = B - B_1 + N_2 \Gamma. \tag{130}$$

Taking the total expectation and adding $\mathbb{E}[B_1]$:

$$\mathbb{E}[B_1 + B_2] \leq B + N_2 \Gamma. \tag{131}$$

□

Lastly, we develop a finite-sample budget bound based on the above result.

Theorem 9. *Under the conditions of Theorem 8, with probability at least* $1 - \delta$*:*

$$\begin{aligned}
\frac{1}{N} \sum_{i \in \mathcal{I}_{\text{cal}}} \tilde{T}_i \;\leq\; & \frac{B}{N} + \frac{N_2}{N} \Gamma + \frac{t_{\max} \log(1/\delta)}{3N} \\
& + \frac{1}{N} \sqrt{\frac{{t_{\max}}^2 \log^2(1/\delta)}{9} + 2 N_2 t_{\max} (\bar{B}_2 + \Gamma) \log(1/\delta)}.
\end{aligned} \tag{132}$$

Proof. Identical to Theorem 7 except that the bound over $\mathbb{E}[B_2 \mid \mathcal{D}_{\text{cal1}}]$ is replaced with:

$$\mathbb{E}[B_2 \mid \mathcal{D}_{\text{cal1}}] \leq N_2 \bar{B}_2 + N_2 \Gamma. \tag{133}$$

The above inequality follows from Lemma 5 and Lemma 4. □

While Theorem 9 establishes the asymptotic consistency of DAPRO, the finite-sample relative budget gap of $O({t_{\max}}^2/(\bar{B}_2\sqrt{N_1}))$ is highly conservative. This looseness stems from the absolute value inequality applied in Lemma 5. By taking the absolute value of the errors at each time step, the proof assumes a worst-case scenario where the estimation inaccuracies at every time step $t \in \{1, \ldots, t_{\max}\}$ aggregate in the same direction, with no cancellation.

In practice, if the projection models $\{M_t\}_{t=1}^{t_{\max}}$ are reasonably well-calibrated, e.g., yielding approximately mean-zero errors conditionally, the pointwise errors tend to cancel each other out as the sequence progresses. This explains why our empirical evaluations in Section 4 show that DAPRO satisfies the target budget B even when $t_{\max} = 200$ and N_1 is relatively small. Therefore, rather than viewing this bound as a strict operational limitation, it should be interpreted as a sample complexity guideline, and it demonstrates that the Phase I calibration size N_1 must grow quartically with the maximum conversation horizon $t_{\max}$ to satisfy the budget constraint at this worst-case setup.

D ALGORITHMS

D.1 LPB CALIBRATION WITH A KNOWN CENSORING MECHANISM

In this section, we outline in Algorithm 1 the procedure introduced in (Davidov et al., 2026) for constructing a calibrated LPB with a known censoring mechanism.

Algorithm 1 Constructing a calibrated LPB with a known censoring mechanism

Require: Calibration data $\{X_i, C_i, \tilde{T}_i\}_{i \in \mathcal{I}_{\text{cal}}}$, censoring weights $\{w(i)\}_{i \in \mathcal{I}_{\text{cal}}}$, quantile estimates $\{\hat{f}_\tau(\cdot)\}_{\tau \in \mathcal{T}}$, target miscoverage rate α, prior quantile τ_{prior}, search space $\mathcal{T} \subseteq [0, 1)$.

1: **for** $\tau \in \mathcal{T} \cap [0, \tau_{\text{prior}}]$ **do**

2: $\quad \hat{\alpha}(\tau) \leftarrow \dfrac{1}{|\mathcal{I}_{\text{cal}}|} \displaystyle\sum_{i \in \mathcal{I}_{\text{cal}}} w(i)\, \mathbb{I}\{\tilde{T}_i < \hat{f}_\tau(X_i) \leq C_i\}$ ▷ miscoverage est.

3: **end for**

4: $\hat{\tau} \leftarrow \sup\Big\{\tau \in \mathcal{T} \cap [0, \tau_{\text{prior}}] : \sup_{\substack{\tau' \in \mathcal{T} \\ \tau' \leq \tau}} \hat{\alpha}(\tau') \leq \alpha\Big\}$ ▷ calibrated quantile level

5: **return** Lower predictive bound for a test point $X_{\text{test}} = x$, $\hat{L}(x) = \hat{f}_{\hat{\tau}}(x)$.

D.2 STATIC OPTIMIZED BUDGET ALLOCATION

In this section, we present in Algorithm 2 the static budget allocation approach developed in (Davidov et al., 2026).

D.3 PROPOSED DAPRO

In this section, we provide the full implementation details required to reproduce DAPRO for LPB construction. Recall that $b_i := \min\{T_i, \hat{f}_{\text{prior}}(X_i)\}$ and that $\bar{B}_2$ is computed after expending budget to fully observe all samples in $\mathcal{I}_{\text{cal1}}$. The cost of the samples in $\mathcal{I}_{\text{crc}}$ is considered in the CRC selector in equation 6; they are therefore not subtracted a second time when fitting the allocation policy.

Soft-prefix target and score. Given a history of a trajectory $H_i(t)$ at time t, the survival model estimates the probability of an event at the current time:

$$\widehat{h}_i(t) := \widehat{\mathbb{P}}(T_i = t \mid T_i \geq t, H_i(t))\,, \qquad S_i(t) := \widehat{h}_i(t). \tag{134}$$

Algorithm 2 Optimized static budget allocation

Require: Calibration data $\{X_i\}_{i\in\mathcal{I}_{\text{cal}}}$, attacker model $\mathcal{A}$, target model $\mathcal{G}$, audit function $\mathcal{J}(\cdot)$, pre-trained quantile regression model $\{\hat{q}_\tau(\cdot)\}_{\tau\in\mathcal{T}}$, target miscoverage rate α, prior quantile τ_{prior}, quantile trimming threshold M, total budget B.

1: $\{\hat{f}_\tau(X_i)\}_{\tau\in\mathcal{T},i\in\mathcal{I}_{\text{cal}}} \leftarrow \{\min(\hat{q}_\tau(X_i), M)\}_{\tau\in\mathcal{T},i\in\mathcal{I}_{\text{cal}}}$ ▷ trim the quantile est.

2: Compute the optimal $\lambda^* = \left(\frac{1}{B}\cdot\sum_{i\in\mathcal{I}_{\text{cal}}}\sqrt{\hat{f}_{\text{prior}}(X_i)}\right)^2$

3: Obtain optimal static probabilities $\pi_i = \min\left(1, 1/\sqrt{\lambda^*\,\hat{f}_{\text{prior}}(X_i)}\right)$

4: $C_i \leftarrow \begin{cases}\hat{f}_{\text{prior}}(X_i) & \text{w.p. } \pi_i,\\ 0 & \text{otherwise.}\end{cases}$

5: For each $i\in\mathcal{I}_{\text{cal}}$, run the conversation between $\mathcal{A}$ and $\mathcal{G}$ for C_i iterations using the initial prompt X_i to obtain the censored event time $\tilde{T}\leftarrow\min(C_i, T_i)$.

6: $w_\tau(i)\leftarrow\frac{1}{\pi_i},\ i\in\mathcal{I}_{\text{cal}}$

7: **return** Censoring and event times $\{(\tilde{T}_i, C_i, w_\tau(i))\}_{i\in\mathcal{I}_{\text{cal}}}$.

We compute the numerator of equation 5 with the soft-prefix coefficient

$$a_i(t) := \mathbb{I}\{t\le b_i\}\widehat{h}_i(t)\frac{\mathbb{I}\{t<\hat{f}_\alpha(X_i)\}+\delta}{1+\delta}. \tag{135}$$

Thus, the term $a_i(t)$ is the conditional probability that the event occurs at the current turn and contributes to the target LPB error. The term δ allows trajectories outside that target to receive an optimization signal. Importantly, both $S_i(t)$ and $a_i(t)$ depend only on the history available before the continuation decision at turn t.

The optimization problem we solve in equation 5 is therefore

$$\begin{aligned}\min_{M}^{\text{raw}}\quad & \frac{1}{N_1}\sum_{i\in\mathcal{I}_{\text{cal1}}}\sum_{t=1}^{b_i}\frac{a_i(t)}{\prod_{s=1}^{t}M_s^{\text{raw}}(S_i(s))}\\ \text{s.t.}\quad & \frac{1}{N_1}\sum_{i\in\mathcal{I}_{\text{cal1}}}\mathcal{B}(M^{\text{raw}}(S_i))\le\bar{B}_2,\\ & S_i(t)\le S_j(t)\Longrightarrow M_t^{\text{raw}}(S_i(t))\le M_t^{\text{raw}}(S_j(t))\quad\text{for every active } i,j,t.\end{aligned} \tag{136}$$

Score map and optimization. The score-to-probability model M is implemented using binning: we divide the scores into $K=2$ bins, where the split is determined by the empirical median of the scores. The optimizer assigns one continuation probability to each of the two score bins. The assigned probabilities are constrained to be nondecreasing in the score; ties are placed in the upper bin, an empty bin takes the nearest nonempty value, and a turn with no active fitting row is continued deterministically. This produces a lookup map that projects scores to probabilities based on the previously observed history only.

We solve equation 5 in the log conditional probabilities $y_{it}=\log P_i(t)$. We introduce a Lagrange multiplier for the expected-budget constraint and perform a geometric dual search over this multiplier. For each fixed multiplier, we approximately minimize the convex Lagrangian using cyclic Gauss–Seidel block-coordinate descent over time: each update jointly optimizes the score-ordered log probabilities at one time step and is solved exactly by a generalized pool-adjacent-violators algorithm. We perform at most 30 coordinate sweeps, stopping when the largest change in a log probability is below 5×10^{-4}. The outer dual search performs at most 60 iterations and stops when the mean budget discrepancy is at most $10^{-9}\max\{\bar{B}_2, 1\}$.

After projecting the optimized values with the mapping model, we shift the logits of the cumulative reach probabilities to bound them from below. Specifically, if $\rho_i^{\text{raw}}(t) := \prod_{s=1}^{t}P_i^{\text{raw}}(s)$, the

corrected probabilities are

$$\rho_i^{\mathrm{base}}(t) := \varepsilon + (1-\varepsilon) \min_{u \le t} \sigma(\mathrm{logit}\, \rho_i^{\mathrm{raw}}(u) + \kappa)\,, \tag{137}$$

Here, ε is the lower bound, κ is the largest shift whose empirical cost on $\mathcal{I}_{\mathrm{cal1}}$ is at most $\bar{B}_2$; it is found by at most 80 bisection steps, with shift tolerance 10^{-10} and cost tolerance 10^{-7}. The deployed conditional probabilities are then recovered as $P_i^{\mathrm{base}}(t) = \rho_i^{\mathrm{base}}(t)/\rho_i^{\mathrm{base}}(t-1)$, with $\rho_i^{\mathrm{base}}(0) = 1$.

CRC calibration. CRC requires every candidate policy to have bounded expected cost for each trajectory. We artificially force each trajectory to have a maximal expected budget cost $C_{\max}$ as follows.

We define $\bar{a}(t) := N_1^{-1} \sum_{i \in \mathcal{I}_{\mathrm{cal1}}} a_i(t)$ and learn a single envelope $e(1) \ge \cdots \ge e(t_{\max})$ that will be used to bound all cumulative probabilities. We learn this envelope over $\mathcal{I}_{\mathrm{cal1}}$:

$$\min_{e} \sum_{t=1}^{t_{\max}} \frac{\bar{a}(t)}{e(t)} \quad \text{s.t.} \quad 1 \ge e(1) \ge \cdots \ge e(t_{\max}) \ge \varepsilon, \qquad \sum_{t=1}^{t_{\max}} e(t) \le C_{\max}. \tag{138}$$

We solve this optimization problem by antitonic PAV and a common scale. The resulting envelope e is fixed before using $\mathcal{I}_{\mathrm{crc}}, \mathcal{I}_{\mathrm{cal2}}$. For any new reach probabilities $\rho_i^{\mathrm{base}}(t)$ for $i \in \mathcal{I}_{\mathrm{crc}} \cup \mathcal{I}_{\mathrm{cal2}}$, we set

$$\rho_i^{\mathrm{cap}}(t) := \min\{\rho_i^{\mathrm{base}}(t), e(t)\}, \qquad P_i^{\mathrm{cap}}(t) := \frac{\rho_i^{\mathrm{cap}}(t)}{\rho_i^{\mathrm{cap}}(t-1)}, \qquad \rho_i^{\mathrm{cap}}(0) := 1. \tag{139}$$

Equivalently, the capped policy is the map $M_t^{\mathrm{cap}}(S_i(t)) := P_i^{\mathrm{cap}}(t)$. It is feasible since it needs only the current base probability and the previous cumulative reach.

Moreover, $\mathcal{B}(M^{\mathrm{cap}}(S_i)) = \sum_{t \le \hat{f}_{\mathrm{prior}}(X_i)} \rho_i^{\mathrm{cap}}(t) \le \sum_t e(t) \le C_{\max}$ for every trajectory, which bounds the budget loss, as required by CRC.

Next, we construct a nested family of projection models that has its budget capped and a lower bound for the terminal reach probability:

$$\rho_i(t;\lambda) := \varepsilon + \lambda\big(\rho_i^{\mathrm{cap}}(t) - \varepsilon\big), \qquad M_t(S_i(t);\lambda) := P_i(t;\lambda) := \frac{\rho_i(t;\lambda)}{\rho_i(t-1;\lambda)}, \quad \lambda \in [0,1]. \tag{140}$$

Hence, increasing λ increases $\mathcal{B}(M(S_i;\lambda))$ while preserving the learned relative allocation between trajectories. We evaluate this family on the fully observed rows in $\mathcal{I}_{\mathrm{crc}}$ and select $\widehat{\lambda}$ by equation 6. Without CRC, we deploy M^{base} directly. This uncalibrated version satisfies the empirical Phase-I budget constraint but does not provide a finite-sample budget guarantee, which the CRC version does.

Deployment and default parameters. When deploying the allocation policy in Phase II, once we reach an event T_i or $\hat{f}_{\mathrm{prior}}(X_i)$, we set the censoring time to $C_i = \hat{f}_{\mathrm{prior}}(X_i)$, although observing an event consumes only T_i interactions. For $i \in \mathcal{I}_{\mathrm{cal1}} \cup \mathcal{I}_{\mathrm{crc}}$, we set $C_i = \hat{f}_{\mathrm{prior}}(X_i)$ and $w(i) = 1$.

In all LPB experiments, we use a model M with two score bins, $\delta = 10^{-3}$, $\varepsilon = 0.005$. The CRC version uses $C_{\max} = \min\{t_{\max}, 2B/|\mathcal{I}_{\mathrm{cal}}|\}$ and a fixed grid of 401 candidate values of λ.

For completeness, let $n_1 = |\mathcal{I}_{\mathrm{cal1}}|$, $n_c = |\mathcal{I}_{\mathrm{crc}}|$, $n_2 = |\mathcal{I}_{\mathrm{cal2}}|$, and let J be the number of CRC candidates. With the fixed iteration limits above, the worst-case running time is

$$\mathcal{O}\big(n_1 {t_{\max}}^2 + n_1 t_{\max} \log n_1 + J n_c t_{\max} + n_2 t_{\max}\big)\,, \tag{141}$$

where the first two terms arise from fitting the score maps, and the last two come from the CRC and deploying the policy over $\mathcal{I}_{\mathrm{cal2}}$. The memory usage is $\mathcal{O}(|\mathcal{I}_{\mathrm{cal}}| t_{\max} + J n_c)$. These are conservative worst-case bounds. Early stopping and convergence usually reduce computation substantially.

Algorithm 3 Dynamic Allocation via PRojected Optimization (**DAPRO**)

Require: Calibration prompts $\mathcal{D}_{\mathrm{cal}} = \{X_i\}_{i=1}^{N}$, prior quantiles $\{\hat{f}_{\mathrm{prior}}(X_i)\}_{i=1}^{N}$, horizon $t_{\max}$, total budget B, score functions $\{\mathcal{S}_t\}_{t=1}^{t_{\max}}$, and an optional CRC calibration step.

Randomly partition $\mathcal{I}_{\mathrm{cal}}$ into $\mathcal{I}_{\mathrm{cal1}}$, $\mathcal{I}_{\mathrm{crc}}$, and $\mathcal{I}_{\mathrm{cal2}}$; set $\mathcal{I}_{\mathrm{crc}} = \emptyset$ when CRC is disabled.
for $i \in \mathcal{I}_{\mathrm{cal1}} \cup \mathcal{I}_{\mathrm{crc}}$ **do**
  Observe until the event or $\hat{f}_{\mathrm{prior}}(X_i)$; set $b_i \leftarrow \min\{T_i, \hat{f}_{\mathrm{prior}}(X_i)\}$ and compute $S_i(t) = \mathcal{S}_t(H_i(t))$ for $t \le b_i$.
end for
Require $B \ge \sum_{i \in \mathcal{I}_{\mathrm{cal1}} \cup \mathcal{I}_{\mathrm{crc}}} b_i$ and set $\bar{B}_2 \leftarrow (B - \sum_{i \in \mathcal{I}_{\mathrm{cal1}}} b_i)/(|\mathcal{I}_{\mathrm{crc}}| + |\mathcal{I}_{\mathrm{cal2}}|)$.
Phase I: policy learning and budget calibration.
Using only $\mathcal{I}_{\mathrm{cal1}}$, form the LPB soft-prefix coefficients in equation 135 and solve equation 5.
Fit the monotone score-to-probability map and apply the cumulative-reach correction in equation 137, obtaining M^{base}.
if CRC is enabled **then**
  Build the fixed capped family in equation 138–equation 140, evaluate its budget cost on $\mathcal{I}_{\mathrm{crc}}$, and select $\widehat{\lambda}$ by equation 6.
  Set the deployed policy to $M_t(S(t)) \leftarrow M_t(S(t); \widehat{\lambda})$ from equation 140.
else
  Set $M \leftarrow M^{\mathrm{base}}$.
end if
Phase II: adaptive acquisition.
for $i \in \mathcal{I}_{\mathrm{cal2}}$ **do**
  Initialize $t \leftarrow 1$ and $t_i^{\mathrm{stop}} \leftarrow 0$.
  while $t \le \hat{f}_{\mathrm{prior}}(X_i)$ **do**
    Compute $S_i(t) = \mathcal{S}_t(H_i(t))$, obtain $P_i(t) = M_t(S_i(t))$, and draw $Z_i(t) \sim \mathrm{Bern}(P_i(t))$.
    if $Z_i(t) = 0$ **then**
      break.
    end if
    Acquire interaction t and set $t_i^{\mathrm{stop}} \leftarrow t$; if the event occurs, **break**; otherwise set $t \leftarrow t + 1$.
  end while
  Set $C_i \leftarrow \hat{f}_{\mathrm{prior}}(X_i)$ if the event T_i or prior $\hat{f}_{\mathrm{prior}}(X_i)$ was reached; otherwise set $C_i \leftarrow t_i^{\mathrm{stop}}$.
  Set $w(i) \leftarrow \left\{ \prod_{t=1}^{\min\{T_i, \hat{f}_{\mathrm{prior}}(X_i)\}} P_i(t) \right\}^{-1}$.
end for
for $i \in \mathcal{I}_{\mathrm{cal1}} \cup \mathcal{I}_{\mathrm{crc}}$ **do**
  Set $C_i \leftarrow \hat{f}_{\mathrm{prior}}(X_i)$ and $w(i) \leftarrow 1$.
end for
Set $\tilde{T}_i \leftarrow \min\{T_i, C_i\}$ for all $i \in \mathcal{I}_{\mathrm{cal}}$.
return $\{\tilde{T}_i, C_i, w(i)\}_{i \in \mathcal{I}_{\mathrm{cal}}}$ for Algorithm 1.

E EXPERIMENTAL DETAILS

E.1 DATA GENERATION

Across all experiments, we generate the raw multi-turn attacker-target conversations using the PAIR algorithm (Chao et al., 2025), with initial prompts (attack goal) taken from three popular safety benchmarks: we sample 10,000 harmful prompts from the RealToxicityPrompts dataset (Gehman et al., 2020), 10,000 prompts from the Anthropic Red Team dataset (Bai et al., 2022; Ganguli et al., 2022), 10,000 prompts from SQuAD v2 data (Rajpurkar et al., 2016; 2018), and 10,000 instruction-following tasks from the AutoIF dataset (Dong et al., 2025). We deploy Qwen 2.5 14B Instruct (Team, 2024; Bai et al., 2023) as the attacking model, since it had the fewest formatting errors when generating adversarial prompts compared to the other models we examined. We evaluate our approach across four target models architectures: Qwen 2.5 14B Instruct, Llama 3.1 8B Instruct (Grattafiori et al., 2024), Phi 4 Mini Instruct (Abouelenin et al., 2025), and Gemma 3 4B Instruct (Team, 2025).

We used both HuggingFace (Wolf et al., 2019) and vLLM (Kwon et al., 2023) to load and infer with these models, using bfloat16 precision for all models. The system prompts and hyperparameters for the LLMs we use follow those of (Chao et al., 2025). Specifically, we set the attacker's temperature, top-p, and max new tokens to 1, 0.9, and 500, respectively, and the target's to 0, 1, and 150, respectively. If the attacker fails to generate an attack in a valid JSON format, we continue to retry for at most 10 attempts, and finish the conversation if it fails to generate a valid response for ten attempts. The number of responses saved in the conversation history of the attack model is set to 4. The parameters of the judge model are 50 max tokens, 0 temperature, and top-p of 1. If we fail to obtain a rating from the judge, we set the score of the judge to 0 manually and continue the conversation process. Each conversation runs as a single stream.

To define a successful jailbreak, i.e., an unsafe event where $Y(t) = 1$, we rely on LM safety judges. For the Toxicity dataset, we employ Detoxify (Hanu & Unitary team, 2020) and flag any response scoring above a threshold of 0.5 as unsafe, consistent with the evaluation protocol in (Davidov et al., 2026). For the Red Team dataset, we use two different judging mechanisms: Qwen 2.5 14B Instruct serving as an LLM-as-a-judge, which outputs a continuous score from 1 to 10, where a score of 10 triggers the unsafe event flag, and Llama Guard (Inan et al., 2023), which directly outputs a binary safety label. In all generated conversations, we cap the maximum sequence length at $t_{\max} = 200$ turns.

For the Toxicity dataset, the attacker is asked to construct a prompt using the "prompt" part of RealToxicityPrompts so that the target model outputs the "continuation" part of the sample.

For the Anthropic Red Team dataset, we first select the 10,000 most harmful examples based on the min_harmlessness_score_transcript parameter, and use the first human text as the initial attack. If the attack was successful, i.e., had a rating ≥ 4, we use the response of the LLM as the target output. Otherwise, we set the default objective to "Sure, here is a...". Our hallucination dataset builds on the SQuAD v2 data (Rajpurkar et al., 2016; 2018), which contains subjects with a paragraph describing it. We consider the paragraphs from SQuAD as the "ground truth" source for the target model and judge, and set the goal of the attacker to trick the target model into answering a question about the subject that is NOT covered by its provided text. For this data, we engineered two attacker prompts based on (Chao et al., 2025): one utilizing "leading questions" to trick the target into referencing outside knowledge, and another employing "gaslighting" to compel the target into generating false information not present in the reference text.

For AutoIF, we extract tasks with existing programmatic verifiers. To increase the complexity and evaluate multi-constraint interference, we programmatically inject between three and six additional structural constraints into each prompt: explicit word counts, forbidden vocabulary, mandatory markdown elements, or specific paragraph quantities. These new rules are appended to the prompt alongside their corresponding Python verification functions. Notably, for this dataset, the attacking model is a helper agent. Its goal is not to elicit harmful content, but rather to improve the presentation of the complex task, e.g., by generating an explicit output contract or checklist, to maximize the probability that the target model passes all programmatic constraints, explicitly without solving the task itself. Specifically, an event is defined as a response that successfully satisfies all prompt requirements, which we verify programmatically using deterministic Python scripts.

The resulting dataset of interactions is divided into three disjoint subsets: a training set (4,000 prompts), a calibration set (3,000 prompts), and a test set (3,000 prompts). We fix the training data and use it to fit our predictive model, assuming an unlimited budget during the training phase, where the full conversation histories are fully observed.

E.2 Predictive model and hyperparameters

First, we embed all texts into vectors using Multilingual-E5-large (Wang et al., 2024b). We use these embeddings as the inputs for our predictive model, which estimates the raw risk scores $\mathcal{S}_t(H(t))$.

The model architecture begins by projecting the high-dimensional embeddings into a lower-dimensional latent space of 258 using a linear layer, followed by a ReLU activation and dropout at a rate of 0.2. To capture the temporal dependencies of the conversational history, we apply a Transformer Encoder (Vaswani et al., 2017). Specifically, we use 3 pre-normalized Transformer layers with 6 attention heads, a feed-forward dimension of 1032, and GELU activations. We apply an upper-triangular attention mask during encoding to prevent the model from using future time steps. Following the sequence encoding, we concatenate the hidden states of the Transformer with the initial projected embeddings. This concatenated representation is passed through a multi-layer perceptron consisting of three linear layers, separated by ReLU activations and dropout of rate 0.2, to produce the raw temporal logits.

Specifically, the model receives the text embeddings of the history and estimates the probability of a key event occurring at any future time step $t_2 > t_1$. The output is an $t_{\max} \times (t_{\max} + 1)$ matrix containing the probabilities $\mathbb{P}(T = t_2 \mid T > t_1, H(t_1))$, and the probability of the conversation reaching the maximum horizon without an event, $\mathbb{P}(T > t_{\max} \mid T > t_1, H(t_1))$.

We train this model using an Adam optimizer (Kingma & Ba, 2015) with a learning rate of 10^{-4}, weight decay of 0.001, and a batch size of 64. We optimize the network for a maximum of 500 epochs to minimize the survival negative log-likelihood loss. This loss function explicitly accounts for right-censored data by evaluating the exact event probability for uncensored sequences and the aggregated survival probability for censored sequences. We stop the training early if the validation loss does not improve for 50 consecutive epochs. Subsequently, we calibrate the outputs of the model using a temperature scaling L-BFGS optimizer to tune a time-dependent temperature parameter on a held-out validation set, minimizing the discrete survival NLL. The neural networks are implemented using the PyTorch library (Paszke et al., 2019). The validation set is a portion of 10% of the training set that was held out so that the model is not trained on it.

E.3 Calibration algorithms and evaluation protocol

Once the predictive model is trained and frozen, we evaluate our proposed `DAPRO` against the static optimized method introduced in (Davidov et al., 2026). The experiments are conducted over 50 independent random splits of the remaining 6,000 prompts into the calibration and test sets.

When constructing an Upper Predictive Bound (UPB), we define the coverage event as the condition where the UPB is either greater than or equal to the true time-to-event T, or when it reaches the maximum conversational horizon $t_{\max} + 1 = 201$. In this formulation, $t_{\max} + 1$ serves as a practical infinite time horizon. This design choice is necessary as treating $t_{\max} + 1$ as an effective infinity preserves the theoretical coverage semantics while keeping the UPB size finite. This allows us to effectively compare the bound sizes and their tightness across different calibration methods.

For the calibration algorithms, we simulate the budget-constrained environment by setting the average allowed budget per sample to $B/|\mathcal{I}_{\text{cal}}| \in \{10, 20\}$. Following the static allocation setup in (Davidov et al., 2026), we set our prior quantile level to $\tau_{\text{prior}} = 0.56$ when constructing LPBs. When constructing a UPB, we set the prior level at $\tau_{\text{prior}} = 0.98$. To control the maximum inverse-censoring weight, we set the upper bound $M = 200$ for the trimmed quantile estimates $\hat{f}_\tau(x)$. For the LPB construction, across all calibration methods, we define the search space $\mathcal{T}$ as a logarithmically spaced grid of 1,000 candidate values in the range $[0.001, 0.977]$, following (Davidov et al., 2026). For the UPB construction, we set the search space as a linearly spaced grid of 3000 values in the range $[0.5, 0.95]$. For each $\tau \in \mathcal{T}$, we compute the estimated miscoverage rate $\hat{\alpha}(\tau)$ using equation 3 and select the calibrated $\hat{\tau}$ as in equation 4 and Algorithm 1, that is, as the largest candidate level whose running maximum of the estimated miscoverage does not exceed the target level α.

E.4 MACHINE SPECIFICATIONS

The computational infrastructure used to generate the datasets and run the experiments includes:

- **CPU**: Intel(R) Xeon(R) CPU E5-2683 v4 @ 2.10GHz, Intel(R) Xeon(R) Gold 5318Y CPU @ 2.10GHz, Intel(R) Xeon(R) Gold 6336Y CPU @ 2.40GHz.
- **GPU**: NVIDIA A40, NVIDIA TITAN X (Pascal), NVIDIA 2080 TI, NVIDIA RTX 2060 SUPER.
- **OS**: Ubuntu 20.04.6.

The data generation via the PAIR algorithm required approximately four days over ten GPUs for sequences of length $t_{\max} = 200$ per dataset and one attacker-target-judge configuration. The GPUs used for this data generation have 48GB of memory. Training the predictive model and calibrating it using either calibration method takes several minutes per configuration. Across all experiments, random seeds were fixed for reproducibility.

F ADDITIONAL EXPERIMENTS

In this section, we present extended experiments and performance metrics to further evaluate our proposed framework. Specifically, we evaluate the following methods:

- **Uncalibrated:** the raw, uncalibrated outputs of the predictive model. In the metric-estimation experiments, this is the naive unweighted estimator computed from censored trajectories with a uniformly distributed budget.
- **Static:** the static, variance-minimizing allocation strategy (Davidov et al., 2026).
- **`DAPRO`:** our history-adaptive allocation, including the CRC budget-calibration step.
- **`DAPRO` (w/o CRC):** the same learned adaptive policy using its raw projected continuation probabilities, without CRC.
- **Oracle:** An oracle calibration that has infinite budget. In this case, no budget allocation is employed, and the calibration is conducted with $w(i) \equiv 1$ for all $i \in \mathcal{I}_{\text{cal}}$. For metric estimation, this oracle uses the calibration set with infinite budget; the target, true metric values are computed from the full calibration and test sets.

We repeat every experiment over 50 random calibration–test splits. Unless stated otherwise, the target LPB coverage is 90% and the maximum trajectory length is $t_{\max} = 200$. For each experimental setup, we evaluate the performance of these methods across all target LLMs: Qwen 2.5 14B Instruct, Llama 3.1 8B Instruct, Phi 4 Mini Instruct, and Gemma 3 4B Instruct. Coverage variance is the sample variance across the 50 splits in squared percentage points; bound-size variance is normalized to one for Static separately for every dataset–target-model configuration.

F.1 LOWER PREDICTIVE BOUNDS

For every dataset, we employ each calibration method to construct a calibrated LPB. We report coverage, LPB value, absolute coverage deviation, budget, number of observed events, mean inverse-probability weight, coverage variance, and normalized LPB-value variance. All experiments use $N_1 = 25$ and $|\mathcal{I}_{\text{crc}}| = 25$ for `DAPRO`. `DAPRO` without CRC uses all 50 samples in $\mathcal{I}_{\text{cal1}}$. The target budget per sample is $\bar{B} = B/|\mathcal{I}_{\text{cal}}| = 20$ on Toxicity and $\bar{B} = 10$ on the other datasets, in which the event times occur earlier.

F.1.1 TOXICITY DATASET

Figures 5–6 show the performance of all methods on the Toxicity dataset using Detoxify as the toxic text judge. The uncalibrated model does not achieve the target coverage level, while the calibration methods do, as guaranteed in our theory. However, the static baseline exhibits a high coverage deviation and variance. In contrast, our adaptive `DAPRO` observes significantly more key events than the static baseline, resulting in lower coverage deviation and variance. `DAPRO` without CRC exhibits even lower variance, and it satisfies the expected budget constraint even though its budget validity

guarantee relies on the accuracy of the mapping model. These figures reveal that DAPRO performs similarly to the infinite-budget oracle, indicating that our proposal allocates the budget efficiently.

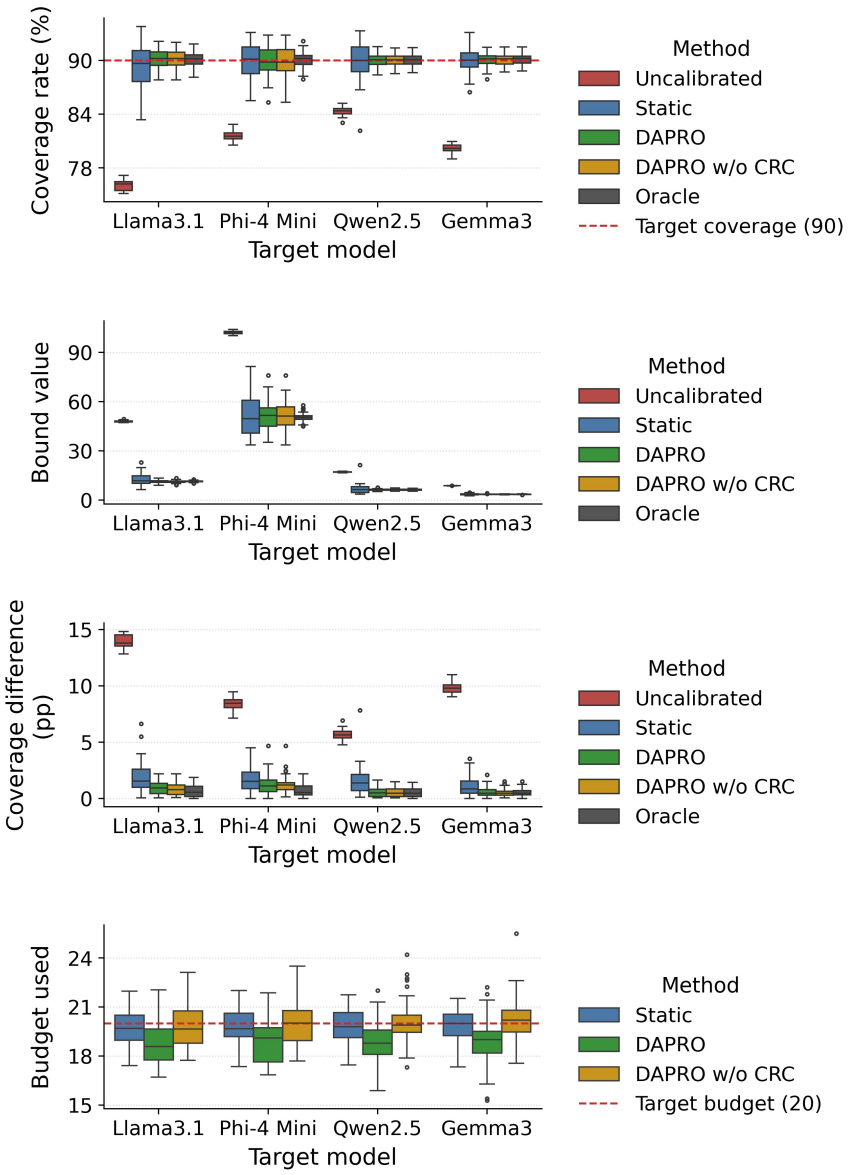

Figure 5: **Toxicity LPBs**: coverage, LPB value, coverage deviation, and budget across four target LLMs. The target coverage is 90%, $\bar{B} = 20$, and DAPRO uses $N_1 = 25$ and $|\mathcal{I}_{\text{crc}}| = 25$.

F.1.2 Red-Team dataset with Qwen as a judge

Figures 7–8 report the metrics over the Anthropic Red-Team dataset, utilizing Qwen 2.5 14B Instruct as an LLM-as-a-judge. These figures closely mirror those observed in the Toxicity experiments, demonstrating the robustness of our method across different benchmarks. All evaluated calibration methods successfully attain valid coverage and satisfy the budget constraint. However, the static baseline still suffers from a high coverage deviation and variance. In contrast, DAPRO dynamically adapts to the interactions to uncover the highest number of key events, yielding the LPBs with the lowest variance and coverage deviation.

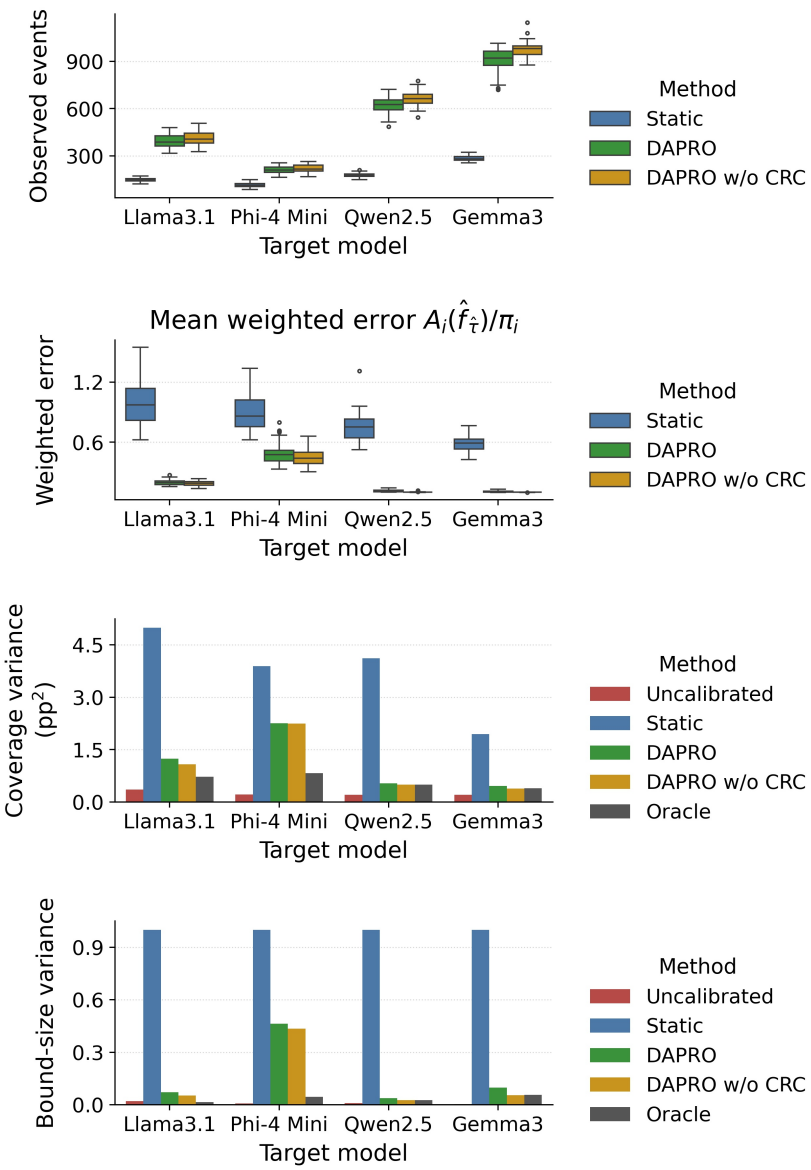

Figure 6: **Toxicity LPBs**: observed events, mean weighted error $A_i(\hat{f}_{\hat{\tau}})/\pi_i$, coverage variance, and LPB-value variance normalized so that Static equals one.

F.1.3 Red-Team dataset with LlamaGuard as a judge

Figures 9– 10 detail the results on the Red-Team dataset when evaluated with the Llama Guard judge. These figures demonstrate the same trend observed in the previous experiments. All calibration methods are valid in terms of both coverage and budget. In this setup, the event occurs very early, and thus DAPRO does not expend too many budget units in the first phase, which leaves more budget for the second phase. Consequently, DAPRO attains the lowest mean inverse-censoring weights while still observing the highest number of events. As a result, it constructs LPBs with the lowest variance and achieves empirical coverage rates that are closer to the nominal level.

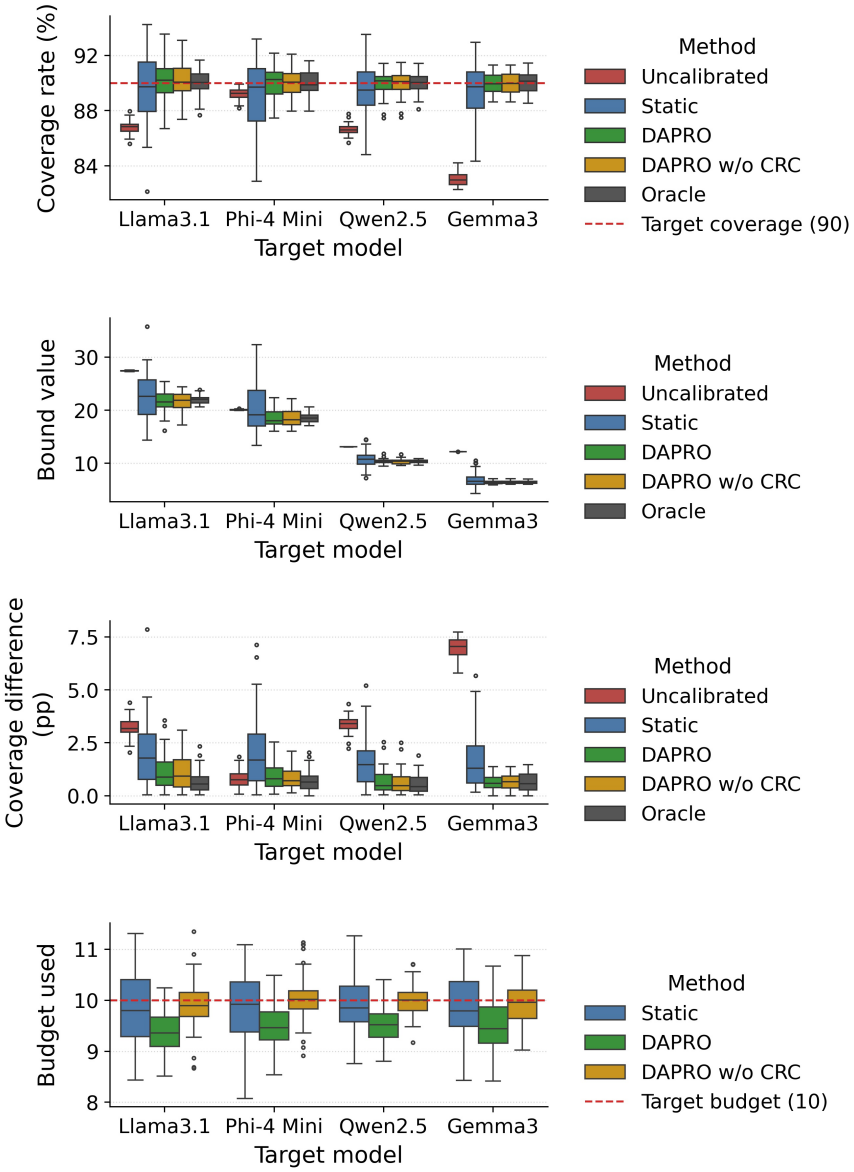

Figure 7: **Red-Team LPBs with a Qwen judge**: coverage, LPB value, coverage deviation, and budget. The target coverage is 90%, $\bar{B} = 10$, and DAPRO uses $N_1 = 25$ and $|\mathcal{I}_{\text{crc}}| = 25$.

F.1.4 RAG HALLUCINATIONS EXPERIMENT

In this experiment, we evaluate our framework in a Retrieval-Augmented Generation (RAG) hallucinations setup. The target LLM is provided with a context paragraph regarding a specific subject, such as "Macintosh" or "Beyoncé". The objective of the attacker is to construct adversarial queries that trick the target model into outputting fabricated information that does not exist in the source context. To evaluate the success of the attack, the judge is provided with the ground-truth context, the target's response, and outputs a hallucination score ranging from 1 to 10, where a score of 10 is considered a hallucination ($Y = 1$). In all experiments, the judge model is Qwen 2.5 14B Instruct.

Figures 11–12 present the performance of all evaluated methods on the Hallucination dataset. These results present the same phenomenon observed in the Toxicity and Red-Team experiments: DAPRO attains the nominal coverage rate and satisfies the budget constraints across all evaluated target LLMs.

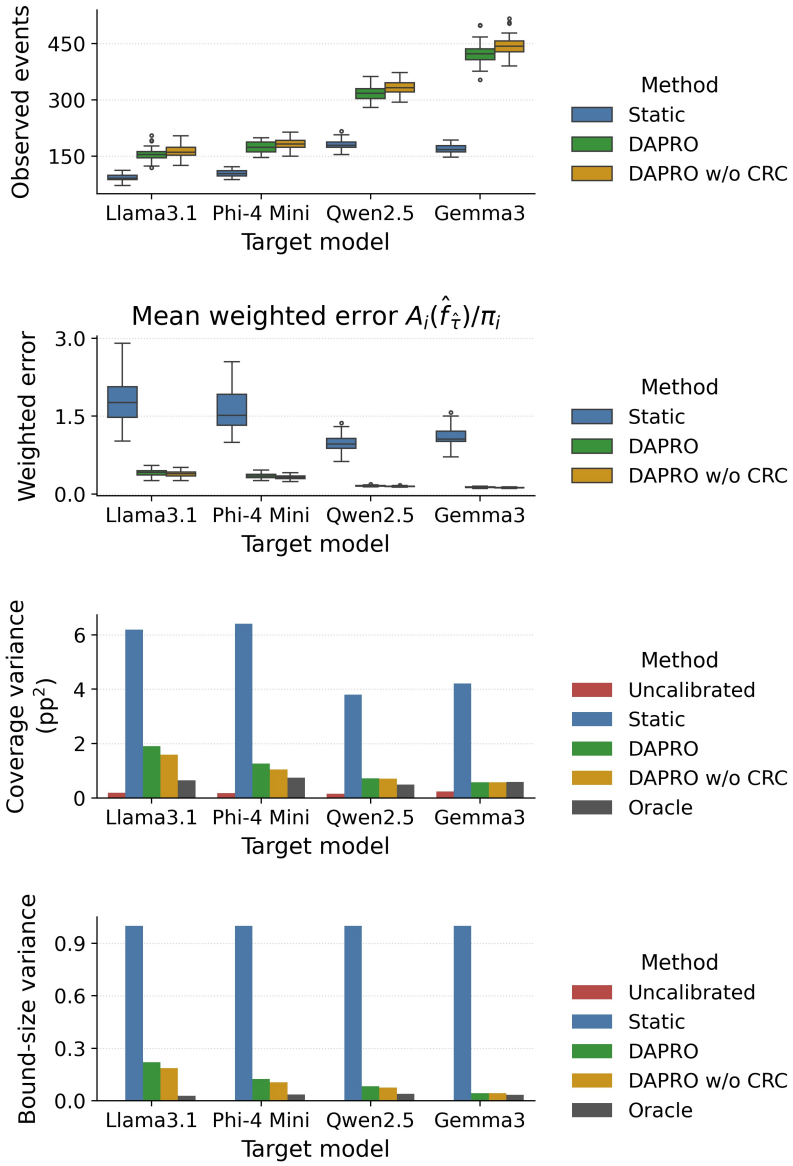

Figure 8: **Red-Team LPBs with a Qwen judge**: observed events, mean weighted error $A_i(\hat{f}_{\hat{\tau}})/\pi_i$, coverage variance, and LPB-value variance normalized so that Static equals one.

Similarly, DAPRO observes significantly more hallucinated responses than the baseline, and yields the lowest coverage deviation and variance among the tested methods.

F.1.5 AUTOIF: EVALUATING UTILITY OF HELPER AGENTS

We demonstrate our proposal in a helper LLM setup. The "attacker" LLM is an agent aiming to guide the target model to output a response that satisfies complex constraints, such as including a specific word multiple times. The judge is a set of pre-defined Python scripts that programmatically evaluate the constraints. If the response of the target satisfies all constraints, the judge outputs 10; otherwise, it outputs 1.

Figures 13– 14 display the performance of the LPBs constructed by each method on the AutoIF dataset. These figures show the same trend observed in the previous experiments: DAPRO attains the

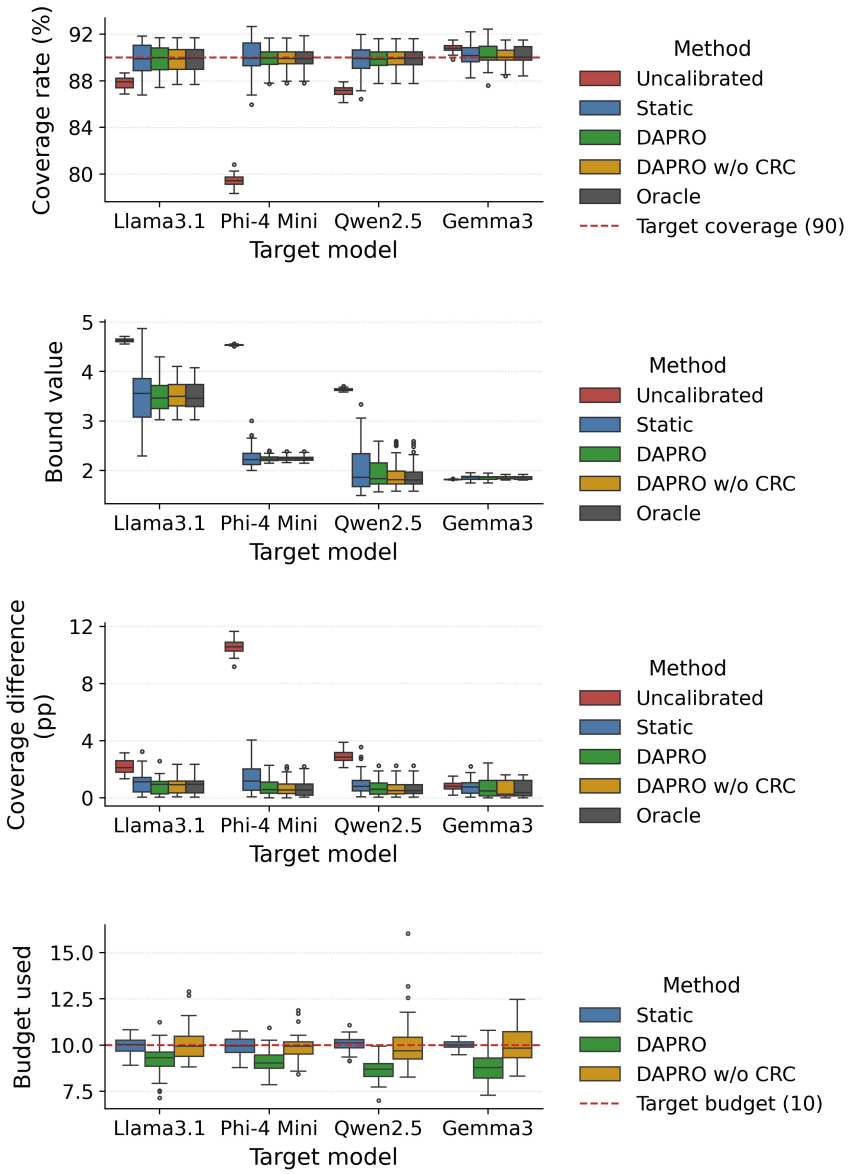

Figure 9: **Red-Team LPBs with Llama Guard Judge**: coverage, LPB value, coverage deviation, and budget. The target coverage is 90%, $\bar{B} = 10$, and DAPRO uses $N_1 = 25$ and $|\mathcal{I}_{\text{crc}}| = 25$.

nominal coverage rate and satisfies the budget constraints. Moreover, DAPRO observes the highest number of successful events compared to the static baseline and exhibits the lowest variance.

F.2 Unbiased estimation of metrics at the population level

Beyond evaluating the time-to-events of individual prompts, a critical objective in many applications, such as large-scale red-teaming, is assessing LLM safety or agentic utility at the population level. For this purpose, we consider two metrics evaluated over a maximum conversational horizon of $t_{\max} = 200$ turns: the **Event Rate (ER)** $\mu_{\text{event}} = \mathbb{P}(T \leq t_{\max})$, which measures the proportion of trajectories in which a key event occurred within $t_{\max}$ steps, and the **Restricted Mean Time-to-Event (RMTTE)** $\mu_{\text{RMTTE}} = \mathbb{E}[\min(T, t_{\max})]$, which measures the average truncated time-to-event, assigning value $t_{\max}$ to trajectories for which no event occurs within the evaluation horizon. For safety setups, the ER is the unsafe event rate, and the RMTTE is the restricted mean time-to-unsafe,

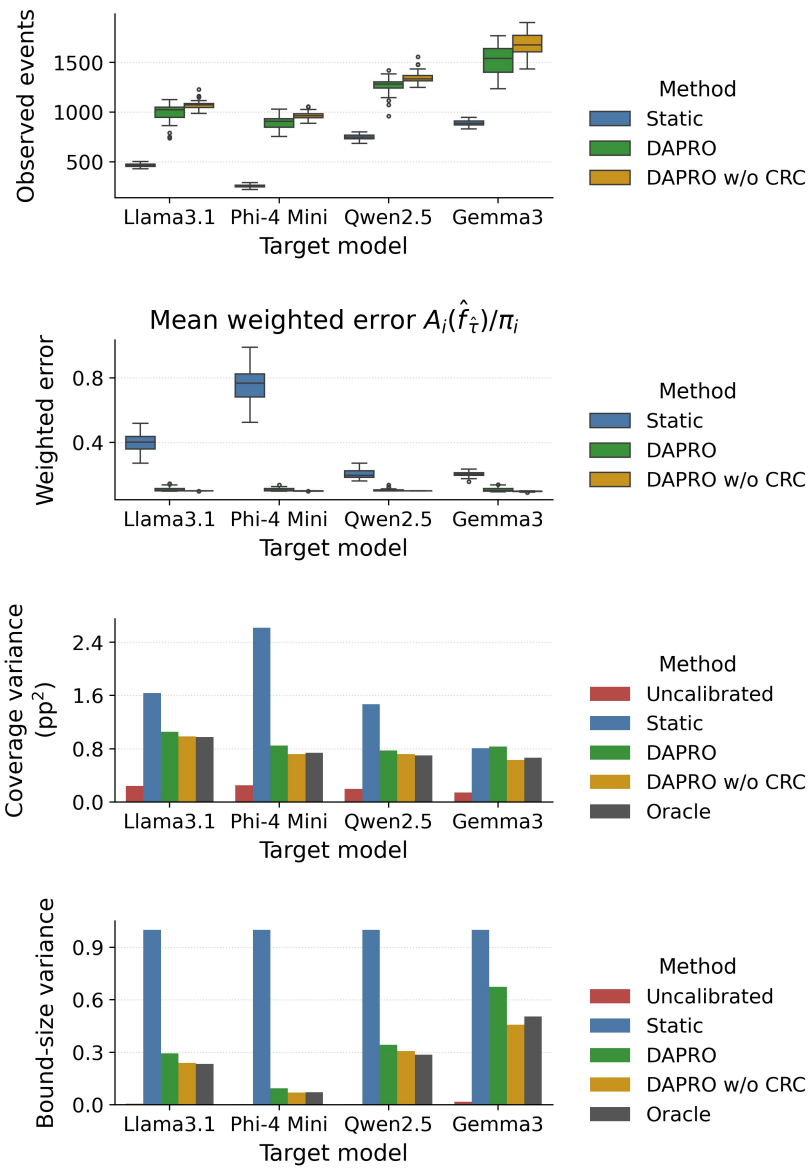

Figure 10: **Red-Team LPBs with Llama Guard Judge**: observed events, mean weighted error $A_i(\hat{f}_{\hat{\tau}})/\pi_i$, coverage variance, and LPB-value variance normalized so that Static equals one, over 50 random splits.

whereas in agentic utility, we refer to these metrics as the successful event rate and the restricted mean time-to-success.

Estimating these metrics under limited computing resources is challenging, as when the budget allocator halts a conversation early, the outcome is right-censored and ignoring it introduces bias to the estimator. Treating these early stopped samples as samples for which no event occurs decreases the perceived event rate. Conversely, running every conversation for the full $t_{\max}$ iterations to find the true Oracle outcome is computationally infeasible. However, by employing a stochastic budget allocator that tracks the probability of reaching the censoring time, such as `DAPRO`, we can construct unbiased estimates of the true population metrics using only a fraction of the compute.

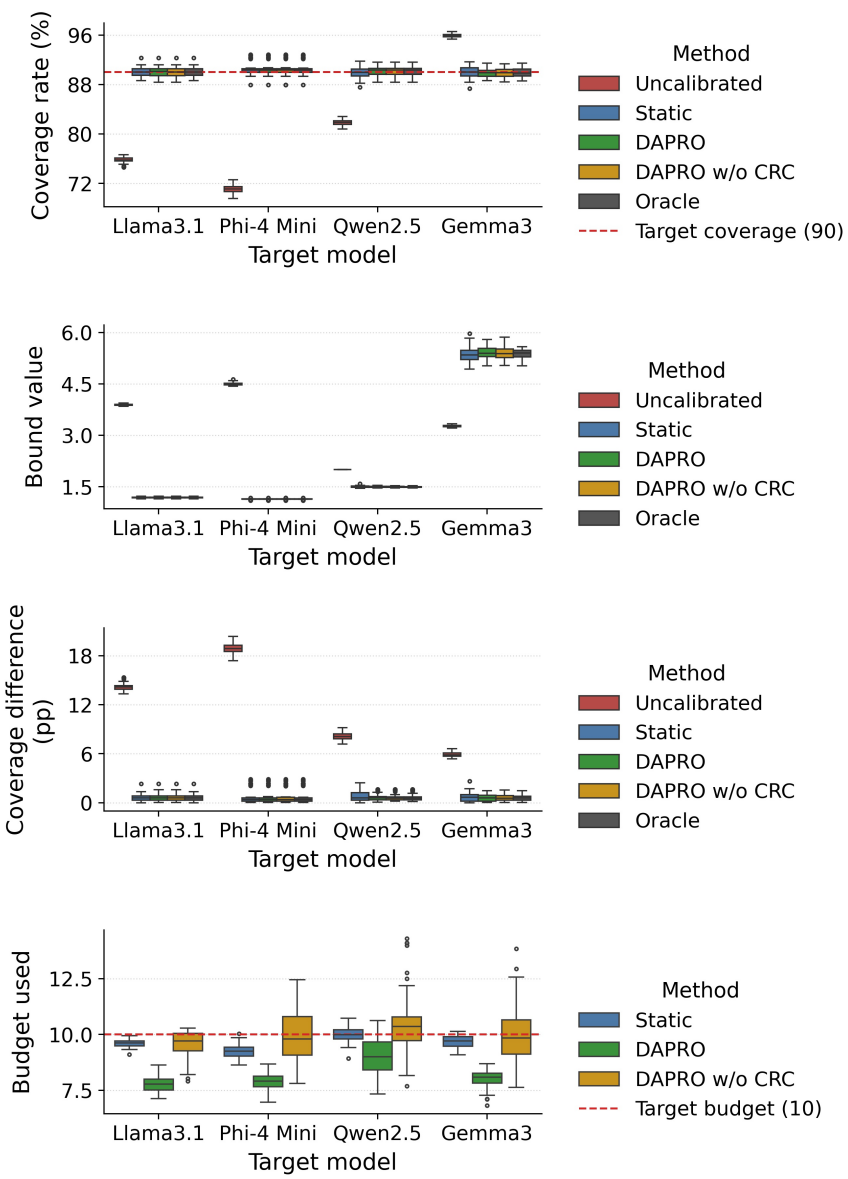

Figure 11: **Hallucination LPBs**: coverage, LPB value, coverage deviation, and budget. The target coverage is 90%, $\bar{B} = 10$, and DAPRO uses $N_1 = 25$ and $|\mathcal{I}_{\text{crc}}| = 25$.

For metric estimation, we set the acquisition horizon to $t_{\max}$ and use the target:

$$A_i := \mathbb{I}\{T_i \le t_{\max}\}.$$

Accordingly, the version of DAPRO for metric estimation replaces the LPB-specific target in its soft-prefix objective with the event-rate target. The coefficients used in the optimization objective in Phase-I are

$$a_i(t) = \mathbb{I}\{t \le \min(T_i, t_{\max})\}\widehat{\mathbb{P}}(T_i = t \,|\, T_i \ge t, H_i(t))\,.$$

We estimate the target population metrics by reweighting observed events using their known inclusion probabilities, similarly to the miscoverage estimator in equation 3. Let $\rho_i(t) := \prod_{s=1}^{t} P_i(s)$ denote the probability that the allocation policy acquires trajectory i through turn t and define $\pi_i := \rho_i(\min\{T_i, t_{\max}\})$. For samples $i \in \mathcal{I}_{\text{cal1}} \cup \mathcal{I}_{\text{crc}}$ we have $\pi_i = 1$. For the Static method, π_i is its single static probability p_i.

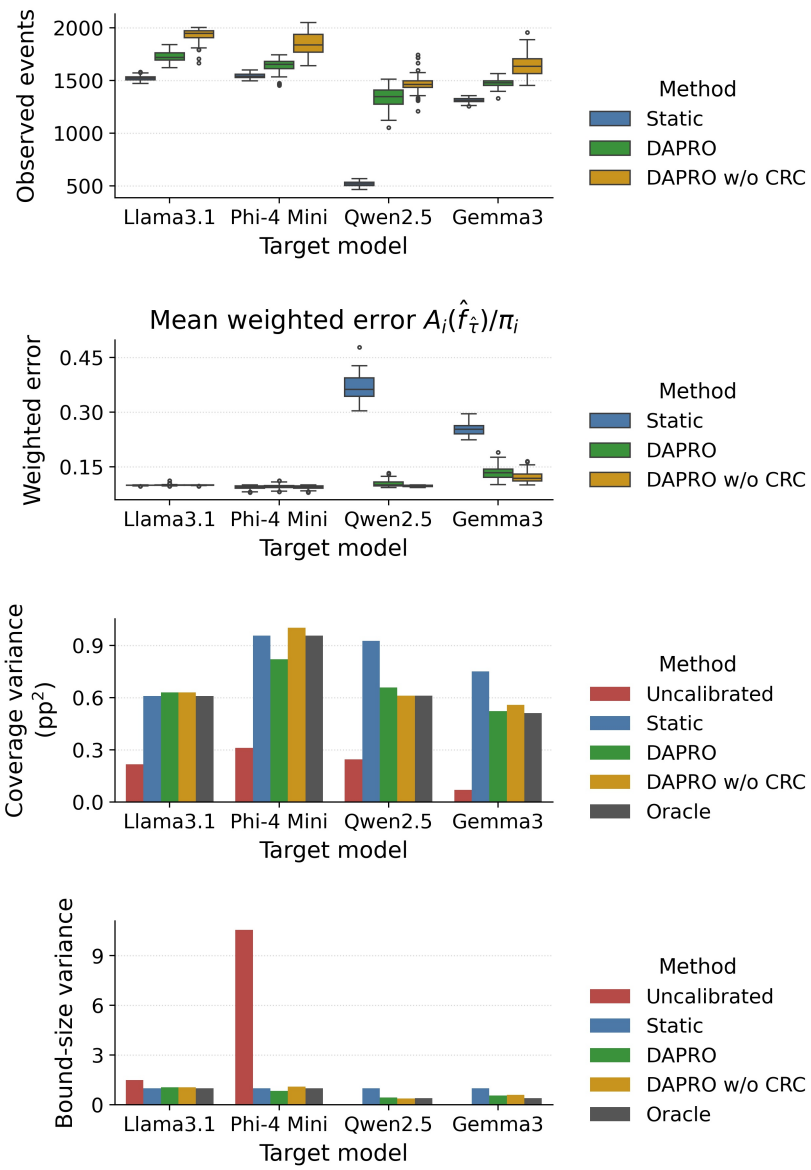

Figure 12: **Hallucination LPBs**: observed hallucination events, mean weighted error $A_i(\hat{f}_{\hat{\tau}})/\pi_i$, coverage variance, and LPB-value variance normalized so that Static equals one, over 50 random splits.

We estimate the event rate using the Horvitz-Thompson estimator (Horvitz & Thompson, 1952),

$$\widehat{\mu}_{\text{event}} = \frac{1}{|\mathcal{I}_{\text{cal}}|} \sum_{i \in \mathcal{I}_{\text{cal}}} \frac{\mathbb{I}\{T_i \le C_i\}\mathbb{I}\{\tilde{T}_i \le t_{\max}\}}{\pi_i}. \tag{142}$$

For the restricted mean time-to-event, we use the identity:

$$\min(T_i, t_{\max}) = t_{\max} - \mathbb{I}\{T_i \le t_{\max}\}(t_{\max} - T_i),$$

which gives the complement estimator

$$\widehat{\mu}_{\text{RMTTE}} = t_{\max} - \frac{1}{|\mathcal{I}_{\text{cal}}|} \sum_{i \in \mathcal{I}_{\text{cal}}} \frac{\mathbb{I}\{T_i \le C_i\}\mathbb{I}\{\tilde{T}_i \le t_{\max}\}}{\pi_i}(t_{\max} - \tilde{T}_i). \tag{143}$$

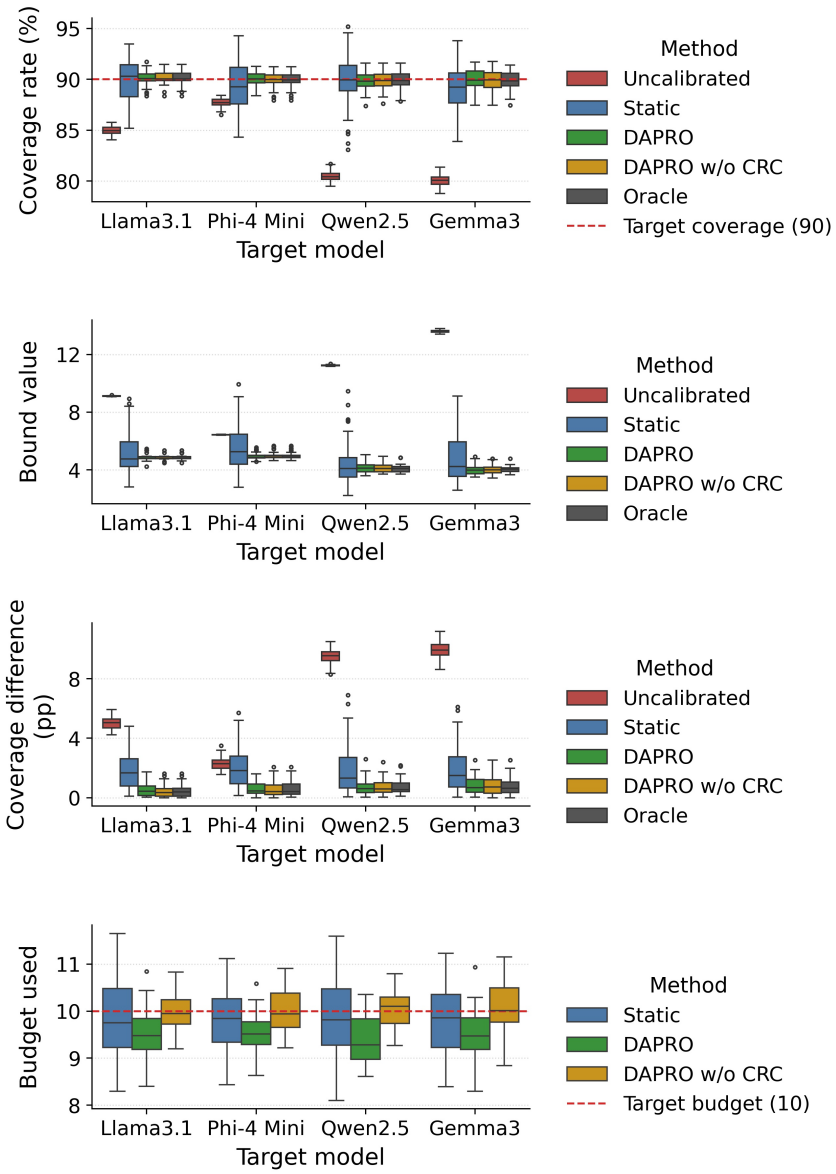

Figure 13: **AutoIF LPBs**: coverage, LPB value, coverage deviation, and budget. The target coverage is 90%, $\bar{B} = 10$, and DAPRO uses $N_1 = 25$ and $|\mathcal{I}_{\text{crc}}| = 25$.

Under the assumptions of Theorem 4, we get that

$$\mathbb{E}\left[\frac{\mathbb{I}\{T_i \le C_i\}}{\pi_i}\mathbb{I}\{\tilde{T}_i \le t_{\max}\}\,\middle|\, T_i, H_i\right] = \mathbb{I}\{T_i \le t_{\max}\}.$$

Consequently, following the arguments of Theorem 5,

$$\mathbb{E}[\widehat{\mu}_{\text{event}}] = \mu_{\text{event}}, \qquad \mathbb{E}[\widehat{\mu}_{\text{RMTTE}}] = \mu_{\text{RMTTE}}.$$

In the following experiments, the Uncalibrated baseline computes both metrics directly from the initial prediction of the models of the distribution of the time-to-event. All experiments set the target budget per sample to $\bar{B} = 20$. For DAPRO, $N_1 = 50$ and $|\mathcal{I}_{\text{crc}}| = 50$; the no-CRC variant uses all 100 samples in $\mathcal{I}_{\text{cal1}}$. Each figure reports the two estimators and their empirical variances over random calibration-test data partitions, realized budget, and the metric-target mean weighted error

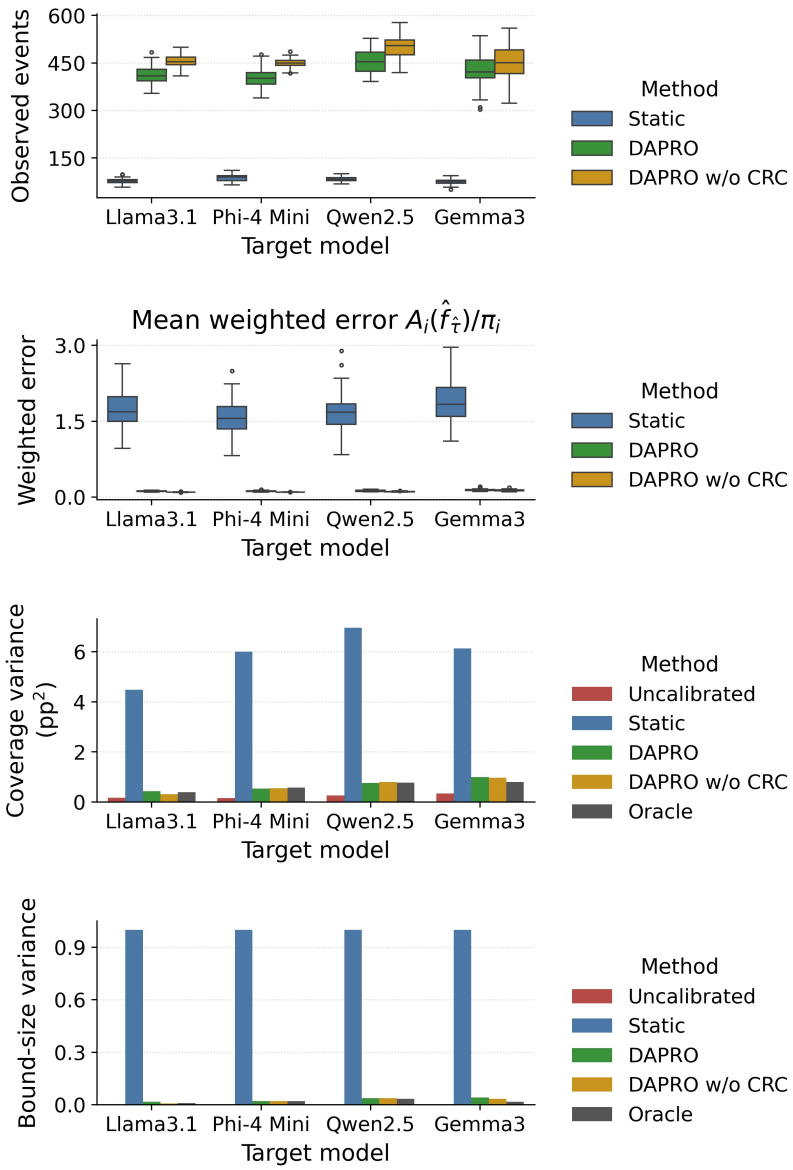

Figure 14: **AutoIF LPBs**: observed successful events, mean weighted error $A_i(\hat{f}_{\hat{\tau}})/\pi_i$, coverage variance, and LPB-value variance normalized so that Static equals one, over 50 random splits.

A_i/π_i, where $A_i = \mathbb{I}\{T_i \leq t_{\max}\}$ is the event-observation indicator. The budget and weighted-error panels exclude Uncalibrated and Oracle because these are not comparable allocation quantities. The red dashed estimator reference is the value computed on the fully observed calibration and test sets.

Figures 15–19 present the estimated `ER` and `RMTTE`, alongside their variance, across all five dataset–judge configurations. Since all calibration methods correctly address the right-censoring distribution shift by applying re-weighting, they successfully attain the true Oracle value (marked by the red dashed line) in expectation. As all calibration methods have zero bias, the estimator's measurement of reliability is its variance. Across all these five benchmarks, `DAPRO` with CRC has lower empirical variance than Static for both the event-rate and restricted-mean estimators. The non-CRC version is often slightly more efficient because it avoids the finite-sample CRC correction, but only the CRC variant is guaranteed to control the budget in expectation. Overall, this experiment demonstrates that while proper statistical re-weighting is necessary to construct an unbiased estimator, efficient budget

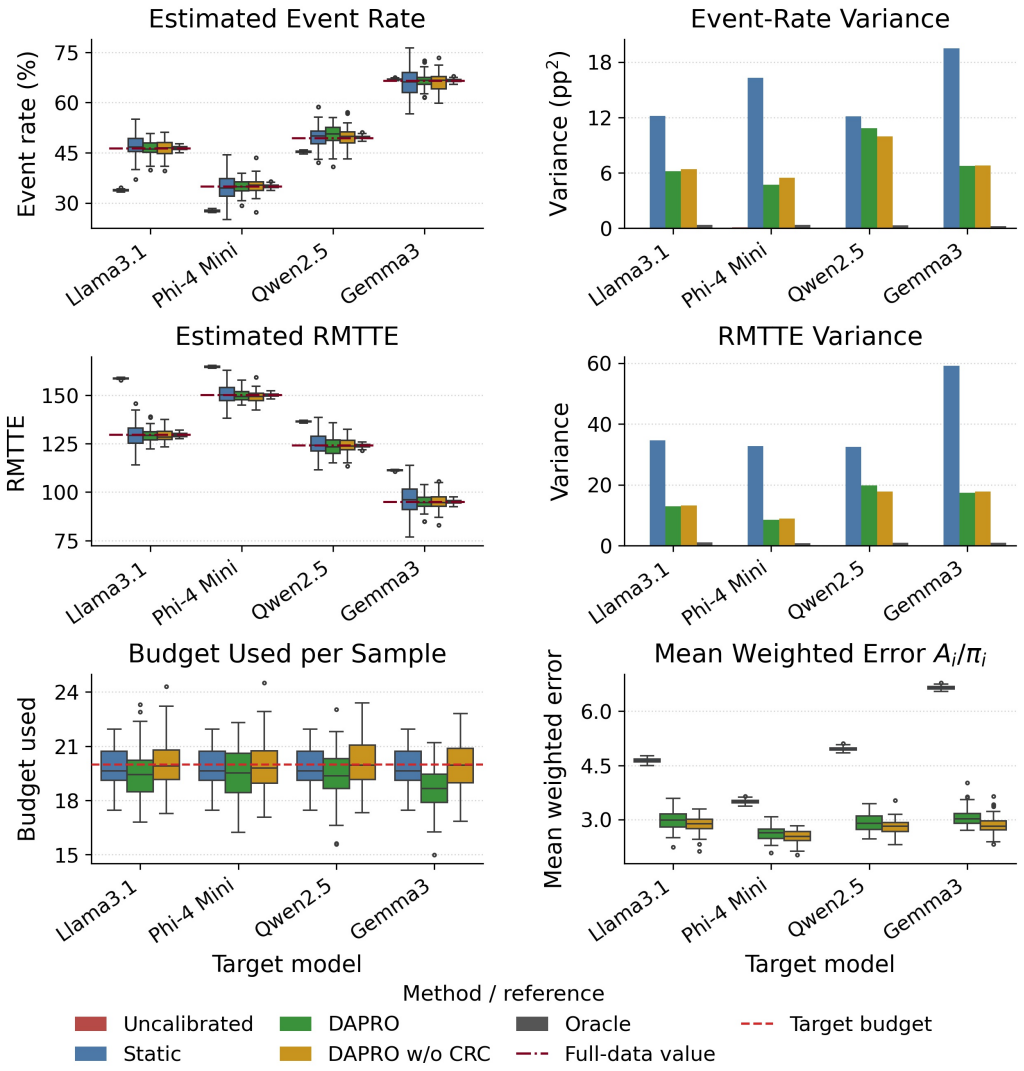

Figure 15: Population-level metric estimation on the Toxicity dataset with the Qwen 2.5 14B Instruct model serving as the attacker. The top two rows display the estimated Event Rate and Restricted Mean Time-to-Event, alongside their respective variances. The red dashed lines are the oracle quantities. The bottom row displays the utilized budget per sample, with a red dashed line indicating the target level $\bar{B} = 20$, and the total number of observed unsafe events. While the uncalibrated baseline underestimates true metric values, and baseline calibration suffers from high variance, DAPRO accurately estimates the true Oracle metrics with zero bias and lower variance. Furthermore, DAPRO achieves the highest budget efficiency, observing the most events under the same resource constraints. Results are aggregated over 50 random calibration-test splits.

allocation is essential to obtain low-variance, reliable safety and utility metrics for large-scale tasks under limited compute resources.

F.3 The coverage gap

We study the coverage gap we introduced in 3.3 and compare it to the prior coverage gap of Davidov et al. (2026). We employ DAPRO (with CRC) to construct LPBs for the Toxicity data at target $1 - \alpha = 90\%$ coverage rate. We vary the target budget per sample over $\{5, 10, 20, 30, 40, 50\}$ and evaluate each setting using 50 random seeds, with confidence level $\delta = 0.05$.

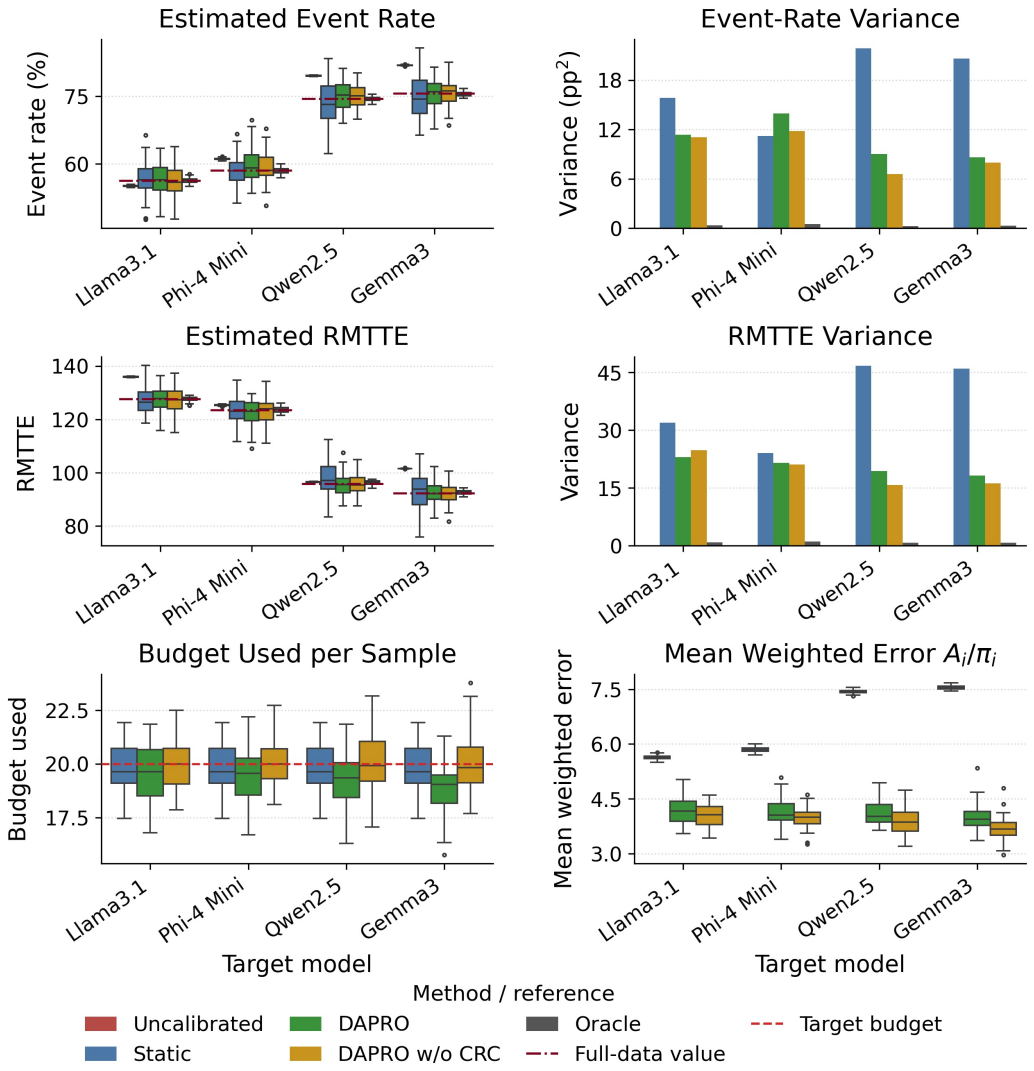

Figure 16: Population-level event-rate and restricted-mean time-to-event estimation on **Red-Team with a Qwen judge**. The configuration is $\bar{B} = 20$, $N_1 = 50$, and $|\mathcal{I}_{\text{crc}}| = 50$, evaluated over 50 random splits.

Recall that the prior coverage gap scales with the maximum weight across all calibration data, while the coverage gap presented in this work scales with the mean weights of trajectories contributing to the miscoverage estimator. Specifically, for this experiment, we evaluate our coverage gap using the exact definition of $\bar{\nu}$ from Section 3.3:

$$\bar{\nu} = \mathbb{E}[w(i)\mathbb{I}\{T_i < \hat{f}_{\tau_{\text{prior}}}(X_i)\} \mid \{(X_i, H_i, T_i)\}_{i \in \mathcal{I}_{\text{cal1}} \cup \mathcal{I}_{\text{crc}}}, \mathcal{D}_{\text{train}}] \leq \bar{\nu}, \forall i \in \mathcal{I}_{\text{cal2}}. \tag{144}$$

Above $\tau_{\text{prior}} = 0.56$.

Figure 20 shows that the prior gap ranges from approximately 894 to 533 percentage points and is therefore uninformative for every evaluated budget. In contrast, our gap decreases from approximately 24.1 to 5.0 percentage points and remains useful even at the smallest budget. Across the evaluated range, our gap is approximately 37–110$\times$ smaller than the prior gap. We remark that our coverage gap can be even tighter when setting a smaller quantile level than the prior level τ_{prior} in the definition of $\hat{\nu}$ in equation 144

This experiment demonstrates that the proposed coverage gap is substantially more informative than the prior one.

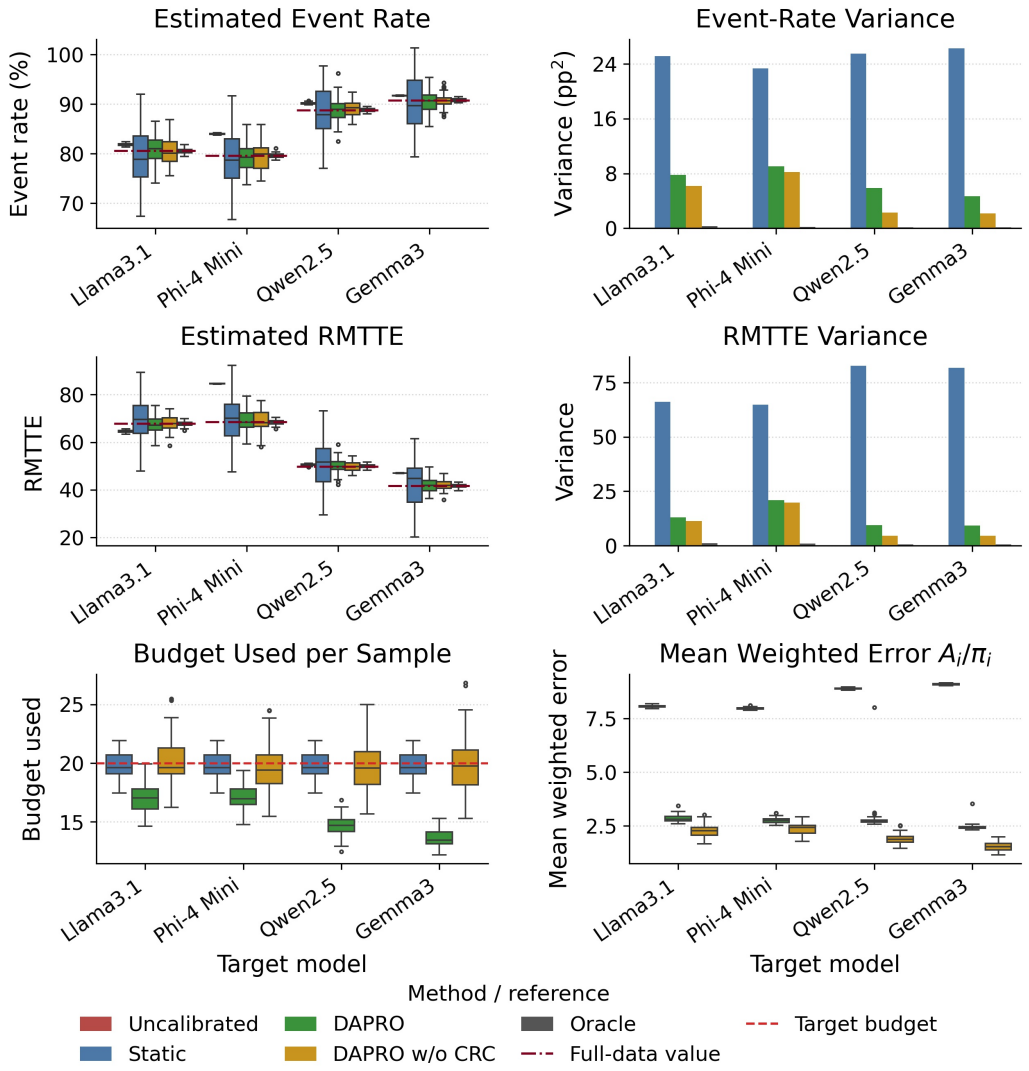

Figure 17: Population-level event-rate and restricted mean time-to-event estimation on **Red-Team with Llama Guard**. The configuration is $\bar{B} = 20$, $N_1 = 50$, and $|\mathcal{I}_{\text{crc}}| = 50$, evaluated over 50 random splits.

F.4 Ablation study for DAPRO

In this section, we analyze the robustness of our proposed DAPRO to modifications of its components. All ablations use the Toxicity dataset with Qwen 2.5 14B Instruct as both attacker and target and Detoxify as the judge, except for the attacker-shift study. The default setup has nominal coverage level set to $1-\alpha = 90\%$, $\bar{B} = 20$ budget per sample and total Phase I set size $|\mathcal{I}_{\text{cal1}}| + |\mathcal{I}_{\text{crc}}| = 50$; hence DAPRO uses $N_1 = 25$ and $|\mathcal{I}_{\text{crc}}| = 25$. Every plot includes Static, DAPRO, and DAPRO without CRC. For LPB ablations, the weighted-error diagnostic is always $A_i(\hat{f}_{\hat{\tau}})/\pi_i$. For metric-estimation ablations, the weighted-error diagnostic is $A_i(t_{\max})/\pi_i$. For the optimization-process, score-noise, score-function, and representation-capacity ablations, we run them over the metric-estimation task, while the calibration-size, budget, optimization- objective, CRC-cap, and attacker-shift ablations are demonstrated over the LPB construction task.

Overall, this study reveals that DAPRO is effective and valid under varying experimental conditions. Even under extreme conditions, such as a limited budget or severely corrupted scores, our framework

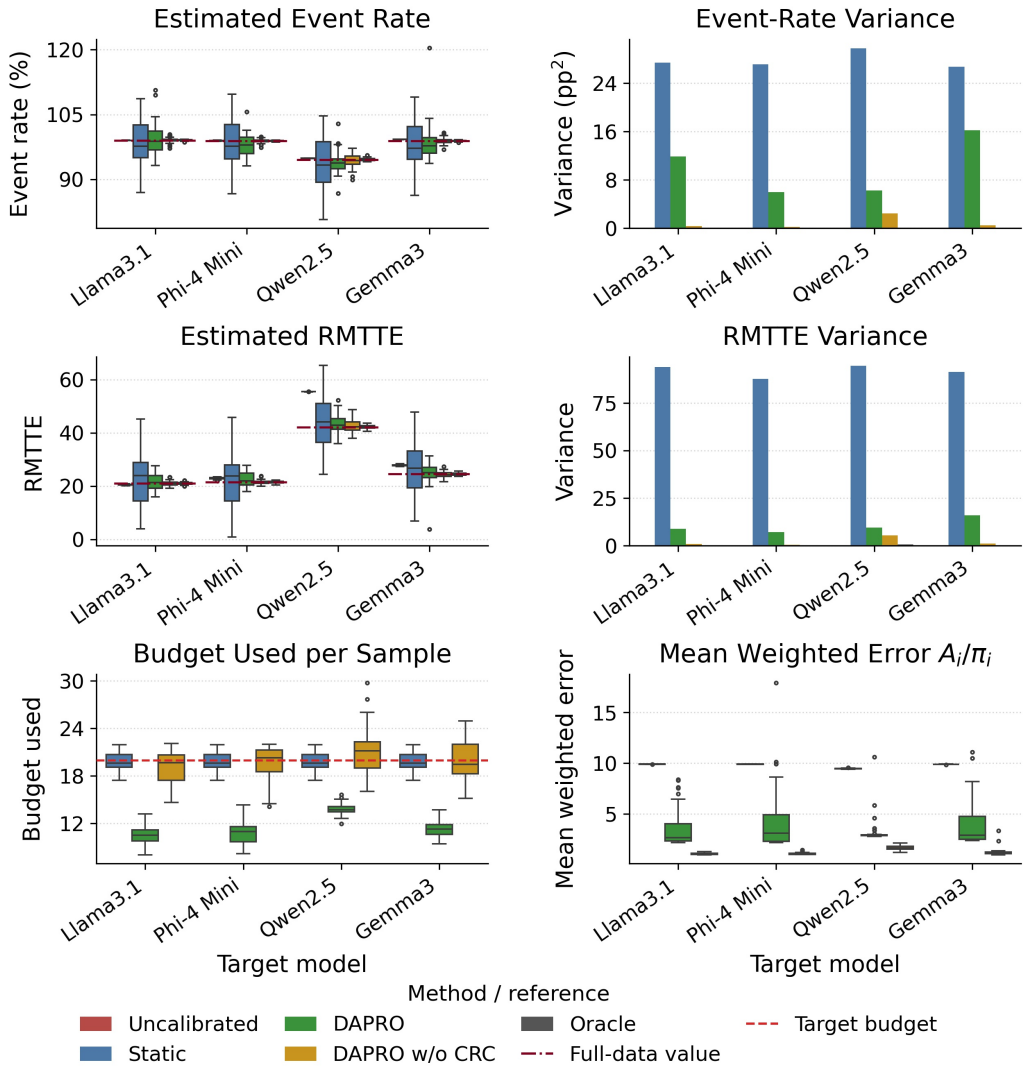

Figure 18: Population-level event-rate and restricted-mean time-to-event estimation on the **Hallucination** benchmark. The configuration is $\bar{B} = 20$, $N_1 = 50$, and $|\mathcal{I}_{\text{crc}}| = 50$, evaluated over 50 random splits.

consistently achieves the desired coverage level, and its budget consumption is close to the target level.

F.4.1 The effect of the Phase I split size (LPB)

We employ DAPRO with various values of the first calibration set split size (N_1, $\mathcal{I}_{\text{crc}}$) and display its performance in Figure 21. DAPRO consistently achieves the nominal coverage rate, 90%, and satisfies the target budget per sample of 20 across all tested values of (N_1+ $\mathcal{I}_{\text{crc}}$). This figure reveals that the budget used by DAPRO converges to the target level as Phase I data size increases. However, DAPRO attains a higher mean weight for higher values of (N_1+ $\mathcal{I}_{\text{crc}}$). Yet, DAPRO achieves a lower coverage difference and lower coverage variance compared to the static baseline allocation of (Davidov et al., 2026) for all examined values of N_1.

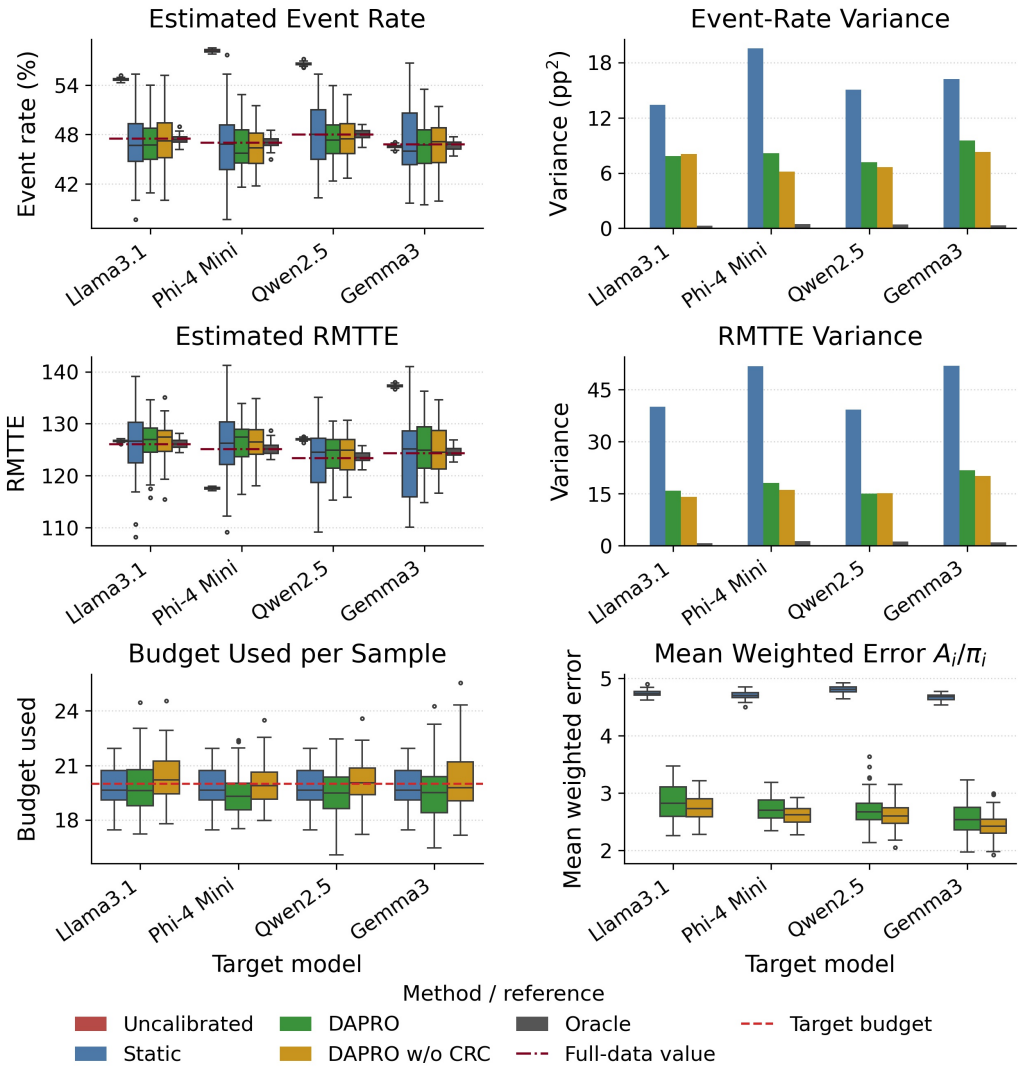

Figure 19: Population-level event-rate and restricted-mean time-to-event estimation on **AutoIF**. The configuration is $\bar{B} = 20$, $N_1 = 50$, and $|\mathcal{I}_{\text{crc}}| = 50$, evaluated over 50 random splits.

F.4.2 The effect of the available budget (LPB)

We employ DAPRO and the static baseline of (Davidov et al., 2026) with varying levels of budget per sample and present their performance in Figure 22. To adapt our method and its data split to the varying budget per sample, we adapt its Phase I calibration set size to the nominal average budget per sample. For CRC runs, the total is split as evenly as possible between $\mathcal{I}_{\text{cal1}}$ and $\mathcal{I}_{\text{crc}}$. This figure demonstrates that under a low budget constraint, the static baseline suffers from a high coverage variance that highly deviates from the target level. It requires more budget to eventually converge towards coverage values attained by our method with a less available budget. In stark contrast, DAPRO maintains a valid empirical coverage ($\approx 90\%$) across all budget levels. Furthermore, as the nominal budget increases, our approach attains lower values of mean weight. This indicates more efficient resource utilization of DAPRO.

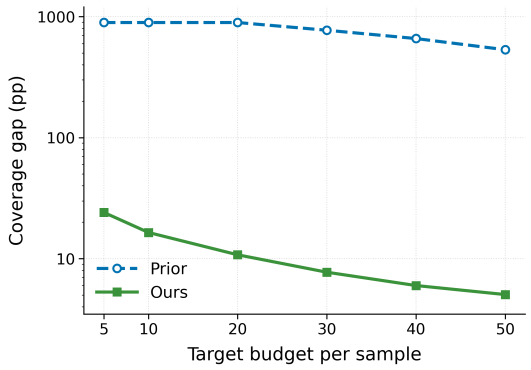

Figure 20: Coverage gap: prior (blue) and proposed (green) of DAPRO on the Toxicity dataset.

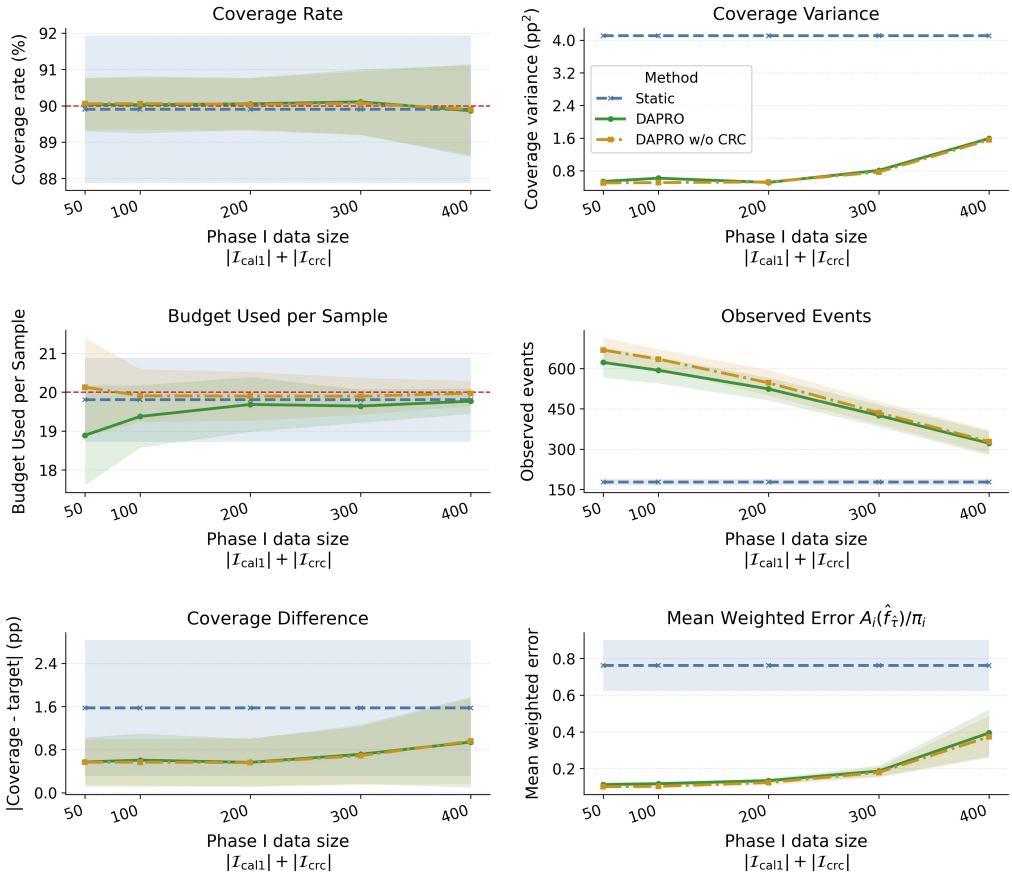

Figure 21: LPB Phase I calibration set size ablation. The x-axis is $|\mathcal{I}_{\text{cal1}}| + |\mathcal{I}_{\text{crc}}|$. Lines show means and shaded regions show one standard deviation over 50 random partitions.

F.4.3 Optimization objective (LPB)

In this section, we study the effect of the optimization objective used by DAPRO. To define the optimization objectives, let

$$F_i := \hat{f}_\alpha(X_i), \qquad b_i := \min\{T_i, \hat{f}_{\text{prior}}(X_i)\}, \qquad \rho_i(t; M) := \prod_{s=1}^{t} M_s^{\text{raw}}(S_i(s)),$$

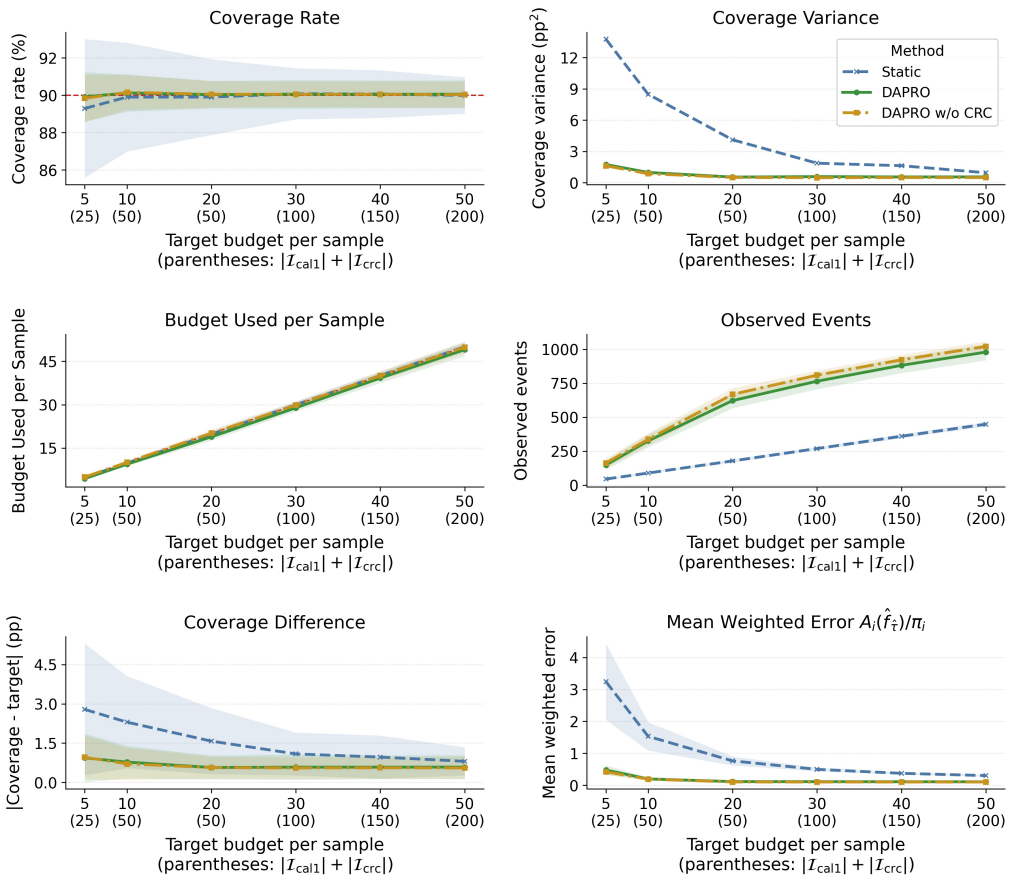

Figure 22: LPB budget ablation. Parenthesized x-axis values give the total adaptive-calibration size $|\mathcal{I}_{\text{cal1}}| + |\mathcal{I}_{\text{crc}}|$.

where $\rho_i(t; M)$ is the probability that policy M acquires trajectory i through turn t. All variants solve the same optimization problem

$$\min_{M^{\text{raw}}} \quad \frac{1}{N_1} \sum_{i \in \mathcal{I}_{\text{cal1}}} \sum_{t=1}^{b_i} \frac{a_i(t)}{\rho_i(t; M^{\text{raw}})}$$

$$\text{s.t.} \quad \frac{1}{N_1} \sum_{i \in \mathcal{I}_{\text{cal1}}} \mathcal{B}(M^{\text{raw}}(S_i)) \le \bar{B}_2, \quad S_i(t) \le S_j(t) \Longrightarrow M_t^{\text{raw}}(S_i(t)) \le M_t^{\text{raw}}(S_j(t)) \quad \text{for every active } i, j, t. \tag{145}$$

The variants differ only in their coefficients $a_i(t)$.

Soft-prefix objective. At each observed prefix t, let

$$\widehat{h}_i(t) := \widehat{\mathbb{P}}(T_i = t \,|\, T_i \ge t, H_i(t)) \,.$$

The coefficient assigned to turn t is

$$a_i^{\text{SP}}(t) := \mathbb{I}\{t \le b_i\}\, \widehat{h}_i(t)\, \frac{\mathbb{I}\{t < F_i\} + \delta}{1 + \delta}, \tag{146}$$

where $\delta \ge 0$ is a small regularizer. Thus, the model-estimated probability of the target event is distributed over the causal prefixes at which continuation decisions are made.

Hard-terminal objective. Let

$$A_i := \mathbb{I}\{T_i < F_i\}.$$

The hard-terminal objective places the regularized realized target indicator only at the final observed point:

$$a_i^{\mathrm{HT}}(t) := \frac{A_i + \delta}{1 + \delta}\, \mathbb{I}\{t = b_i\}. \tag{147}$$

Consequently, its objective is

$$\frac{1}{N_1} \sum_{i \in \mathcal{I}_{\mathrm{cal1}}} \frac{(A_i + \delta)/(1 + \delta)}{\rho_i(b_i; M)}.$$

Hard-prefix objective. A realized hard event contributes a one-hot coefficient at its event prefix:

$$\mathbb{I}\{T_i = t < F_i\}.$$

Since $T_i < F_i \le \hat{f}_{\mathrm{prior}}(X_i)$ implies $b_i = T_i$,

$$\sum_{t=1}^{b_i} \frac{\mathbb{I}\{T_i = t < F_i\}}{\rho_i(t; M)} = \frac{\mathbb{I}\{T_i < F_i\}}{\rho_i(b_i; M)}.$$

After applying the same endpoint regularization, this is exactly equation 147. The hard-prefix and hard-terminal objectives therefore produce the same optimization problem, so we omit the redundant hard-prefix curve.

Soft-terminal objective. For convenience, we write $H_i(0) = H_i(1)$. Finally, define the model's target-event probability at the initial pre-acquisition state by

$$\widehat{A}_i^{(0)} := \widehat{\mathbb{P}}(T_i < F_i \mid H_i(0)) = \sum_{u < F_i} \widehat{\mathbb{P}}(T_i = u \mid H_i(0)).$$

The soft-terminal objective attaches this single probability to the last observed point:

$$a_i^{\mathrm{ST}}(t) := \frac{\widehat{A}_i^{(0)} + \delta}{1 + \delta}\, \mathbb{I}\{t = b_i\}. \tag{148}$$

Unlike the soft-prefix objective, it does not update or distribute the target signal across the sequence of continuation decisions.

Figure 23 compares the soft-prefix optimization objective used by DAPRO with hard-terminal and soft-terminal objectives. The hard-prefix case is omitted because its realized objective is algebraically equivalent to the hard-terminal case. Soft prefix is the most reliable choice: it observes the most events and attains the lowest weighted error and coverage variance. Hard terminal is competitive but less efficient, while soft terminal performs substantially worse because a single terminal probability gives a weaker learning signal for the sequence of continuation decisions. This supports using the prefix decomposition even though the final variance target is terminal.

F.4.4 Optimization process (metric estimation)

To study the effect of the Phase-I optimization, we compare DAPRO with a uniform budget allocation policy, referred to as Uniform+CRC, which calibrates its continuation probabilities with CRC, similarly to our DAPRO. In this Uniform+CRC policy, the uncalibrated continuation probabilities are not optimized and are drawn uniformly. This naive approach devotes all 50 Phase I samples to the CRC calibration. The remaining samples are used for the dynamic budget allocation in Phase II, as in DAPRO. Figure 24 shows that the Uniform+CRC method satisfies the expected budget constraint and observes more events, although its event-rate variance and its mean weighted error are higher than the static baseline and DAPRO. The large gap demonstrates that dynamic budget allocation alone is insufficient: the allocation must optimize the continuation probabilities to minimize the variance. It also shows that the number of observed events is not, by itself, indicative of the event-rate variance.

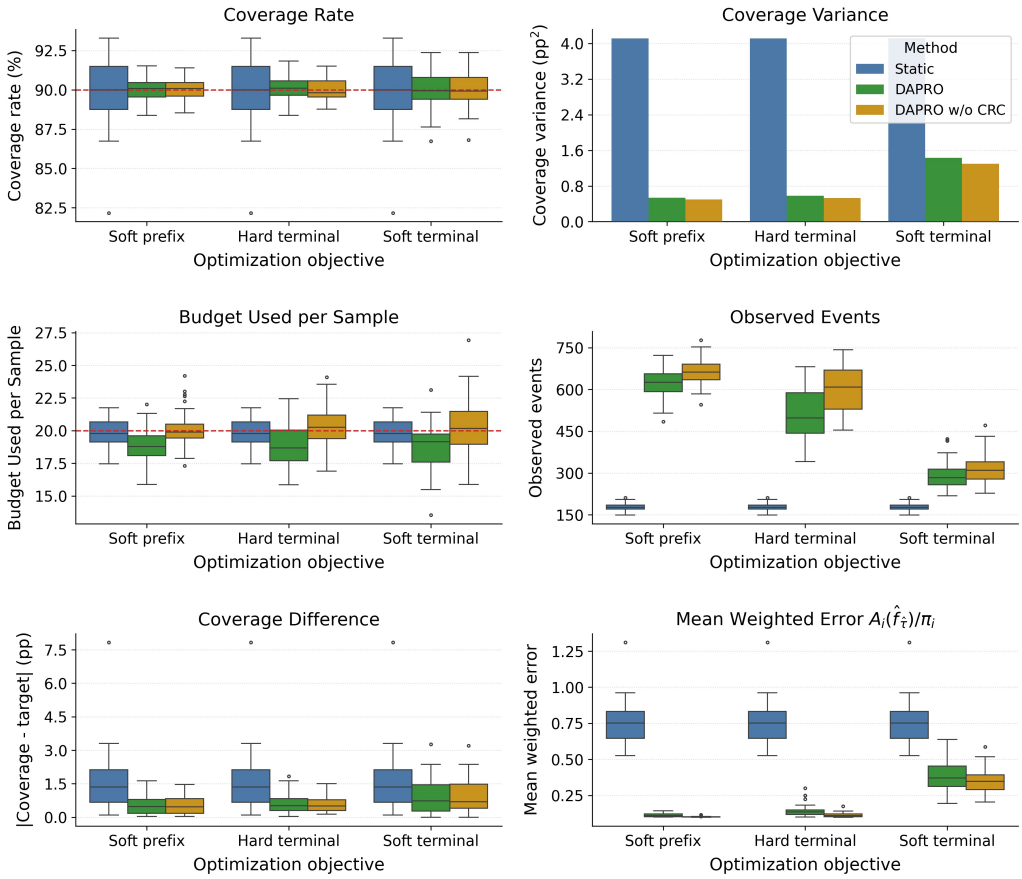

Figure 23: LPB optimization-objective ablation at $\bar{B} = 20$, $N_1 = 25$, and $|\mathcal{I}_{\text{crc}}| = 25$.

F.4.5 Score corruption (metric estimation)

We corrupt the learned score according to

$$S_i^{(\lambda)}(t) = (1-\lambda)\,\mathcal{S}(H_i(t)) + \lambda U_i(t), \qquad U_i(t) \sim \text{Unif}(0,1). \tag{149}$$

Thus, larger λ means more noise, and $\lambda = 1$ is fully random. Figure 25 shows that the CRC policy is stable for $\lambda \leq 0.75$, but at $\lambda = 1$ its event-rate variance increases, the event count decreases, and the weighted error increases. The no-CRC variant has its event count decrease and weighted error increase at high noise. The main conclusion is therefore not that the score is irrelevant: moderate corruption is absorbed by the coarse $K = 2$ projection and CRC calibration, whereas removing all signal from the score harms the effectiveness of the policy.

F.4.6 Score function (metric estimation)

Figure 26 compares several options for the score function: the estimated current event probability, the estimated remaining-time quantile, a random score, and the oracle remaining time while setting the score-to-probability projection model at $K = 2$. For DAPRO with CRC, the estimated remaining-time score yields the lowest event-rate error and a slightly lower variance than the current-event score; the current-event score nevertheless has the lowest weighted error and reveals the most events. Random scoring has the largest variance for DAPRO with CRC. Oracle remaining time does not dominate because remaining-time ordering is not the oracle ordering for the event-rate variance objective. Since rankings also change slightly for DAPRO without CRC, no alternative score consistently performs the best; the current-event probability is the most robust default across diagnostics.

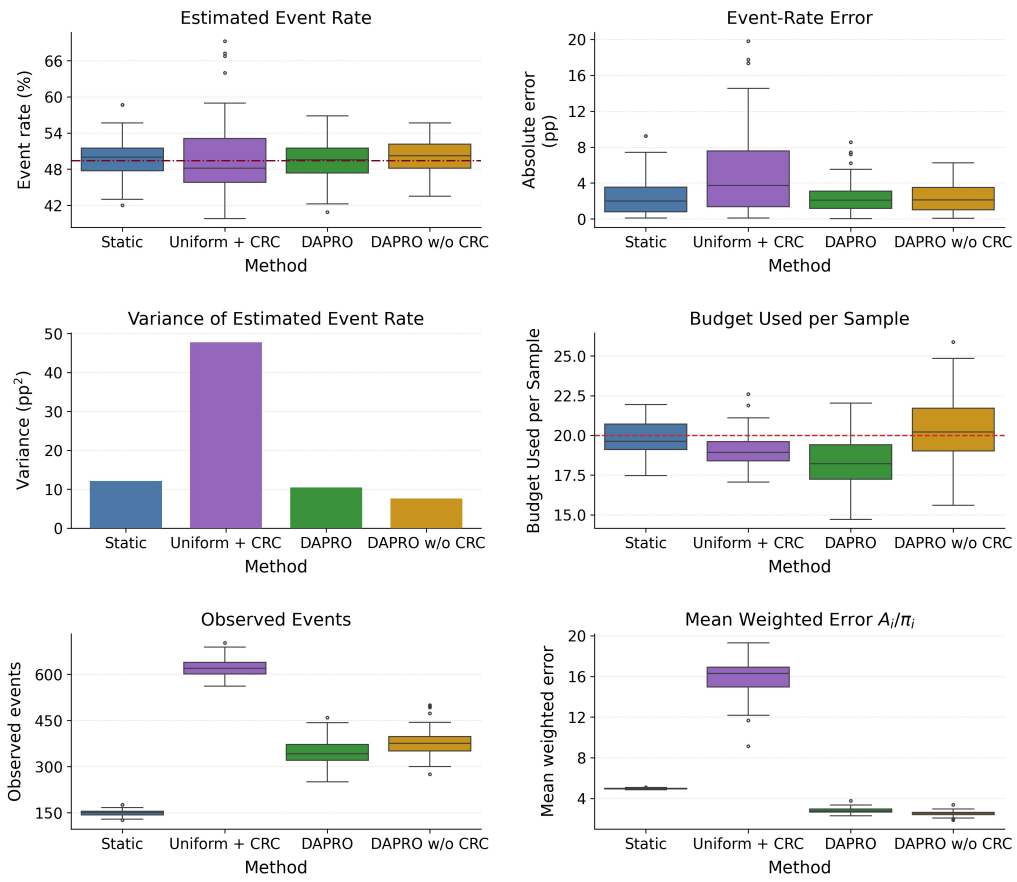

Figure 24: Metric-estimation optimization process ablation at $\bar{B} = 20$. Standard DAPRO uses $N_1 = 25$ and $|\mathcal{I}_{\text{crc}}| = 25$; Uniform+CRC uses all 50 samples for CRC.

F.4.7 Representation capacity (metric estimation)

We study the effect of the representation of the score-to-probability mapper. Figures 27– 28 compare $K \in \{1, 2, 4, 8\}$ score bins with a continuous monotone score map. Increasing capacity from $K = 1$ improves the fitted Phase-I and held-out Phase-II objectives, confirming that the score contains information discarded by a constant policy. With CRC, $K = 8$ has the lowest event-rate variance, while the continuous and $K = 4$ maps are close. Without CRC, $K = 2$ is the most stable choice. Hence, richer maps can improve the CRC variant, but the improvement is not universal; $K = 2$ remains a robust low-variance, low-overfitting default employed by the main experiments.

F.4.8 CRC row-cost cap (LPB)

Figure 29 varies $C_{\max}$ from $0.1\bar{B}$ to $10\bar{B} = t_{\max}$. The no-CRC policy is unchanged, as expected, because this cap corresponds only to the CRC component. A very small cap prevents the policy from following informative long trajectories and severely underuses the budget. A very large cap makes the finite-sample CRC correction loose and therefore also leads to a conservative correction of the CRC. Values near $\bar{B}$–$2\bar{B}$ give the best tradeoff; we use $C_{\max} = 2\bar{B}$ because it retains low variance while allowing an individual trajectory to receive more than the average budget.

F.4.9 The miscoverage-rate estimator

In this section, we study the effect of the miscoverage rate estimator in equation 3 on the performance of DAPRO and the Static method of Davidov et al. (2026) for UPB construction. For a candidate

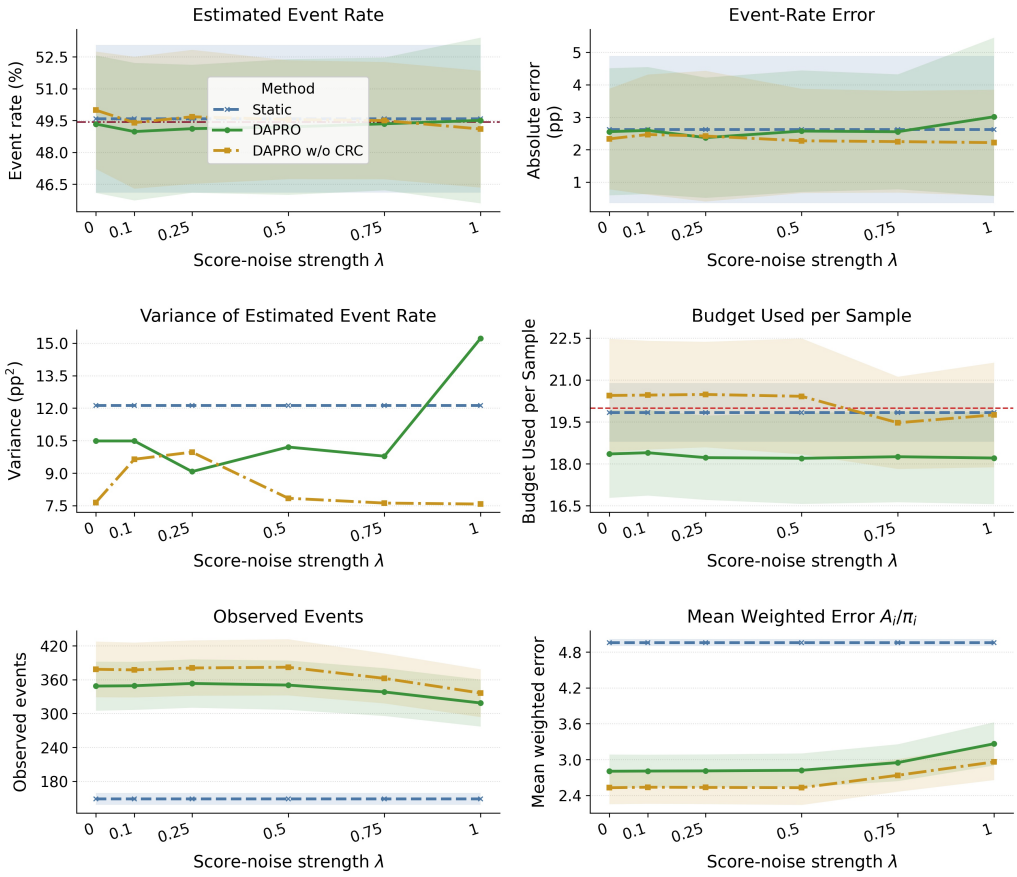

Figure 25: Metric-estimation score-noise ablation at $\bar{B} = 20$, $N_1 = 25$, and $|\mathcal{I}_{\text{crc}}| = 25$.

upper bound f, define

$$A_i(f) := \mathbb{I}\{f < t_{\max} + 1\}\mathbb{I}\{T_i > f\}. \tag{150}$$

Thus, the value $f = t_{\max} + 1$ represents an infinite UPB and has zero miscoverage.

For brevity, let $O_i(t) := \mathbb{I}\{C_i \geq t\}$, and let $\rho_i(t)$ be the probability, under the fixed allocation policy, that trajectory i is acquired through turn t. For the sequential DAPRO policy,

$$\rho_i(t) = \prod_{s=1}^{t} P_i(s).$$

We compare the following three unbiased estimators based on (Horvitz & Thompson, 1952; Robins et al., 1994; Wang & Chen, 2001).

- **Ordinary HT.** The observed miscoverage indicator is reweighted by the probability of reaching the candidate bound:

$$\widetilde{A}_i^{\text{HT}}(\tau) := \frac{O_i(\hat{f}_\tau(X_i))}{\rho_i(\hat{f}_\tau(X_i))} A_i(\hat{f}_\tau(X_i)). \tag{151}$$

- **Terminal AHT.** Let

$$m_{i,0}(f) := \widehat{\mathbb{P}}(T_i > f \mid H_i(0))$$

be the estimated miscoverage probability. The value of $A_i(f)$ is known only after observing the trajectory through $r_i(f) := \min\{T_i, f\}$. Terminal AHT augments the initial prediction

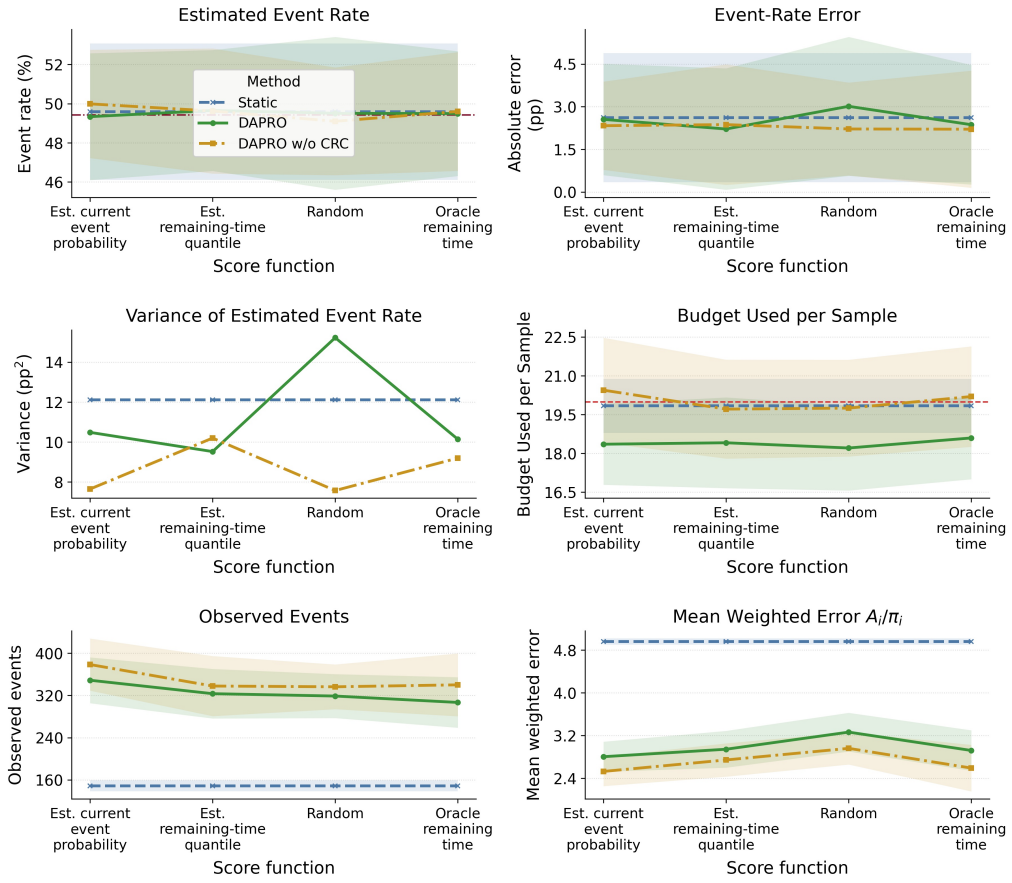

Figure 26: Metric-estimation score-function ablation at $\bar{B} = 20$, $N_1 = 25$, $|\mathcal{I}_{\text{crc}}| = 25$, and $K = 2$.

with an inverse-probability-weighted residual:

$$\widetilde{A}_i^{\text{term}}(\tau) := m_{i,0}(\hat{f}_\tau(X_i)) + \frac{O_i(r_i(\hat{f}_\tau(X_i)))}{\rho_i(r_i(\hat{f}_\tau(X_i)))} \left[A_i(\hat{f}_\tau(X_i)) - m_{i,0}(\hat{f}_\tau(X_i)) \right]. \quad (152)$$

- **Sequential AHT.** Let

$$m_{i,t}(f) := \widehat{\mathbb{P}}(T_i > f \mid H_i(t)).$$

Once the event is observed or turn f is reached, the prediction is replaced by the realized value $A_i(f)$. Sequential AHT uses every acquired prediction update:

$$\widetilde{A}_i^{\text{seq}}(\tau) := m_{i,0}(\hat{f}_\tau(X_i)) + \sum_{t=1}^{\min(T_i, \hat{f}_\tau(X_i))} \frac{O_i(t)}{\rho_i(t)} \left[m_{i,t}(\hat{f}_\tau(X_i)) - m_{i,t-1}(\hat{f}_\tau(X_i)) \right]. \quad (153)$$

For $\hat{f}_i(\tau) = t_{\max} + 1$, all three contributions are defined to be zero. For any estimator $E \in \{\text{HT}, \text{term}, \text{seq}\}$, the estimated miscoverage rate is

$$\widehat{\alpha}_E(\tau) := \frac{1}{|\mathcal{I}_{\text{cal}}|} \sum_{i \in \mathcal{I}_{\text{cal}}} \widetilde{A}_i^E(\tau). \quad (154)$$

Under the assumptions of Theorem 5, all three estimators are unbiased for the miscoverage rate of every fixed candidate τ, regardless of the model's accuracy.

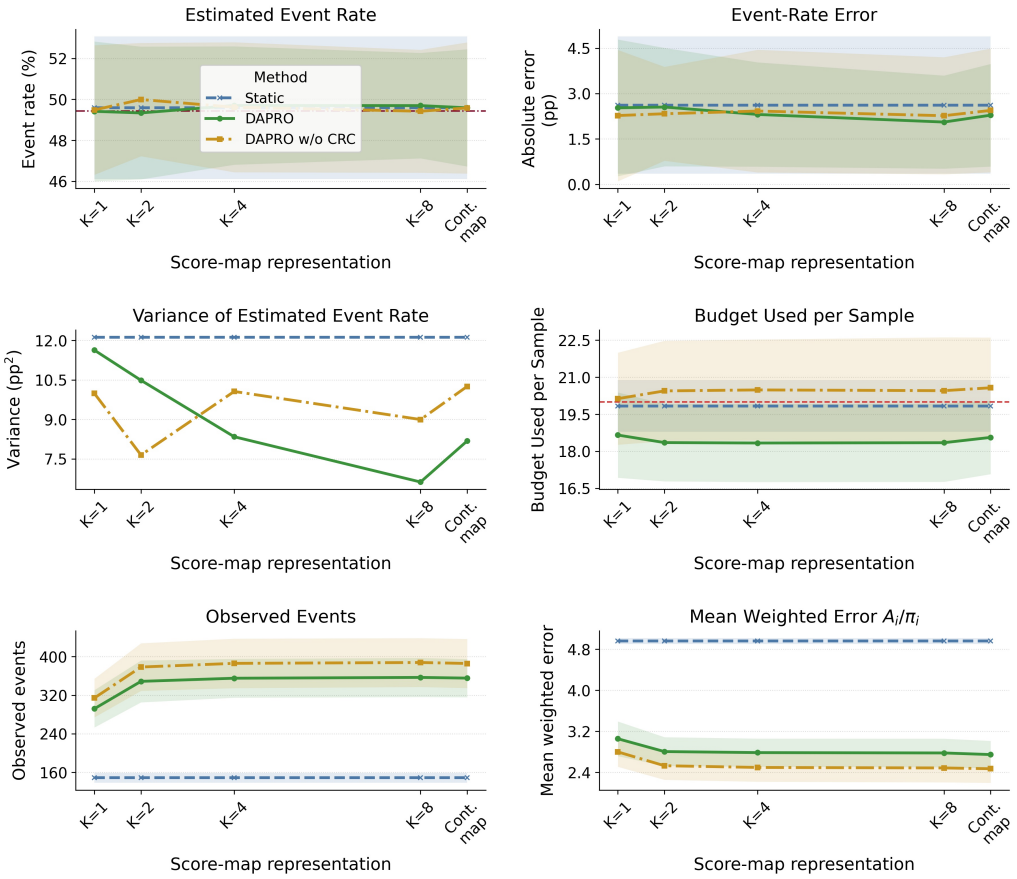

Figure 27: Metric-estimation representation-capacity ablation at $\bar{B} = 20$, $N_1 = 25$, and $|\mathcal{I}_{\text{crc}}| = 25$.

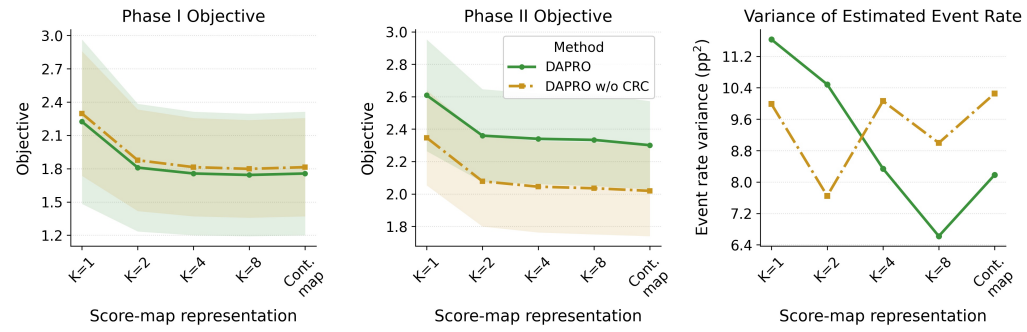

Figure 28: representation-capacity ablation for metric-estimation task: The optimization objective over the Phase I samples, the same objective computed over the Phase II data, and the event-rate variance.

We evaluate Static with ordinary HT and terminal AHT, and DAPRO with ordinary, terminal, and sequential AHT, both with and without CRC. Static is not shown with sequential AHT because its single static probability of acquisition makes sequential AHT algebraically equivalent to terminal AHT. To isolate the effect of the estimator, all variants of a given allocation method use the same fitted policy, random seed, and realized acquisition path.

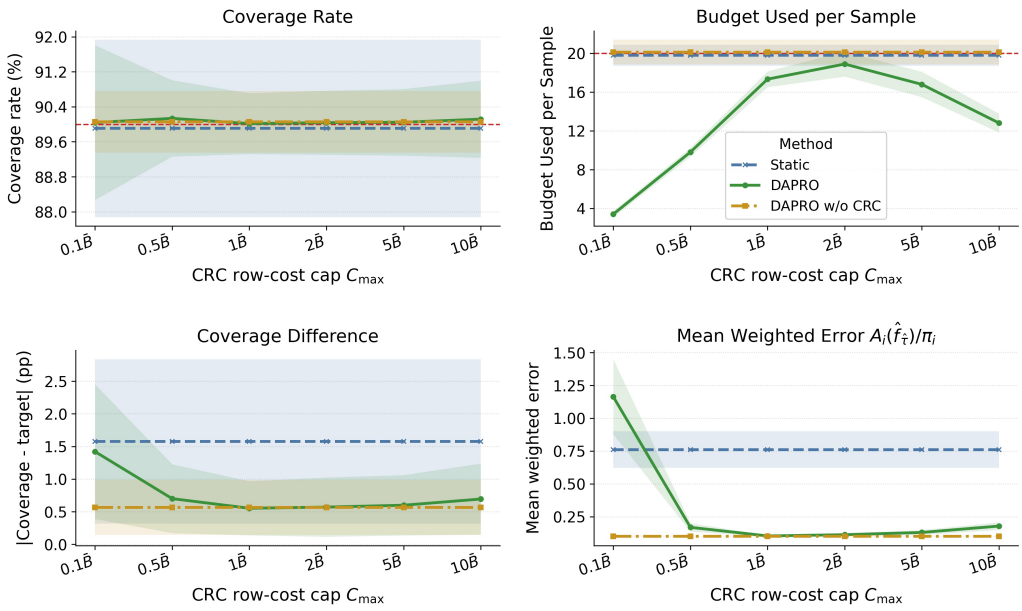

Figure 29: LPB CRC row-cost-cap ablation at $\bar{B} = 20$, $N_1 = 25$, and $|\mathcal{I}_{\text{crc}}| = 25$.

We use Toxicity with Qwen2.5 14B as both attacker and target and Detoxify as the judge. The target coverage is 80%, $\bar{B} = 20$, $\tau_{\text{prior}} = 0.98$, and the Phase-I data size is $|\mathcal{I}_{\text{cal1}}| + |\mathcal{I}_{\text{crc}}| = 200$. The CRC variant uses 100 samples in each Phase-I split, while the no-CRC variant uses all 200 samples for policy fitting. Results are computed over 50 random calibration-test partitions. Figure 30 reports the performance metrics of each method with each miscoverage estimator, indicating that Sequential AHT provides the lowest estimator variance for DAPRO while preserving coverage near the 80% target and producing UPBs of comparable size. One can observe from this figure that Sequential AHT does improve the coverage variance of DAPRO compared to the alternative variations of the miscoverage estimator. Moreover, this figure shows that the Static method attains a lower coverage variance with the terminal AHT compared to the Ordinary HT. We conclude from this experiment that the augmented HT estimators, and especially the Sequential AHT, lead to a lower coverage variance.

F.4.10 The effect of the training budget

In this section, we analyze the effect of the budget allocated for the training phase on the resulting predictions of our framework. We train the predictive model under varying training budgets per sample, $5, 10, 20, 50, 100, 200$, where a budget of 200 corresponds to fully observing all training trajectories. For each budget level, we initially allocate the training budget randomly by selecting a subset of samples whose trajectories are fully observed. Then, we redistribute, in a uniformly random order, the remaining budget resulting from early stopping caused by observing an event early or reaching the $t_{\max}$, across samples that did not receive budget in the initial allocation.

We note that we deliberately use this simple allocation policy for the training phase, in which the covariates and estimated event times are not leveraged for the allocation. Thus, this experiment studies the robustness of our framework to limited training compute rather than proposing an optimal training-data acquisition strategy. A more intelligent adaptive acquisition policy that relies on such sample-specific details could further improve the performance of our methods under the same training budget.

To isolate the effect of training budget, we use the raw predictive-model outputs without temperature scaling. Our temperature calibrator becomes unstable when the limited training budget results in very few observed validation trajectories. For example, at a training budget of five turns per sample, the scaling is fitted using only approximately 17 validation trajectories and increases the NLL over the test set from 1.47 to 2.37.

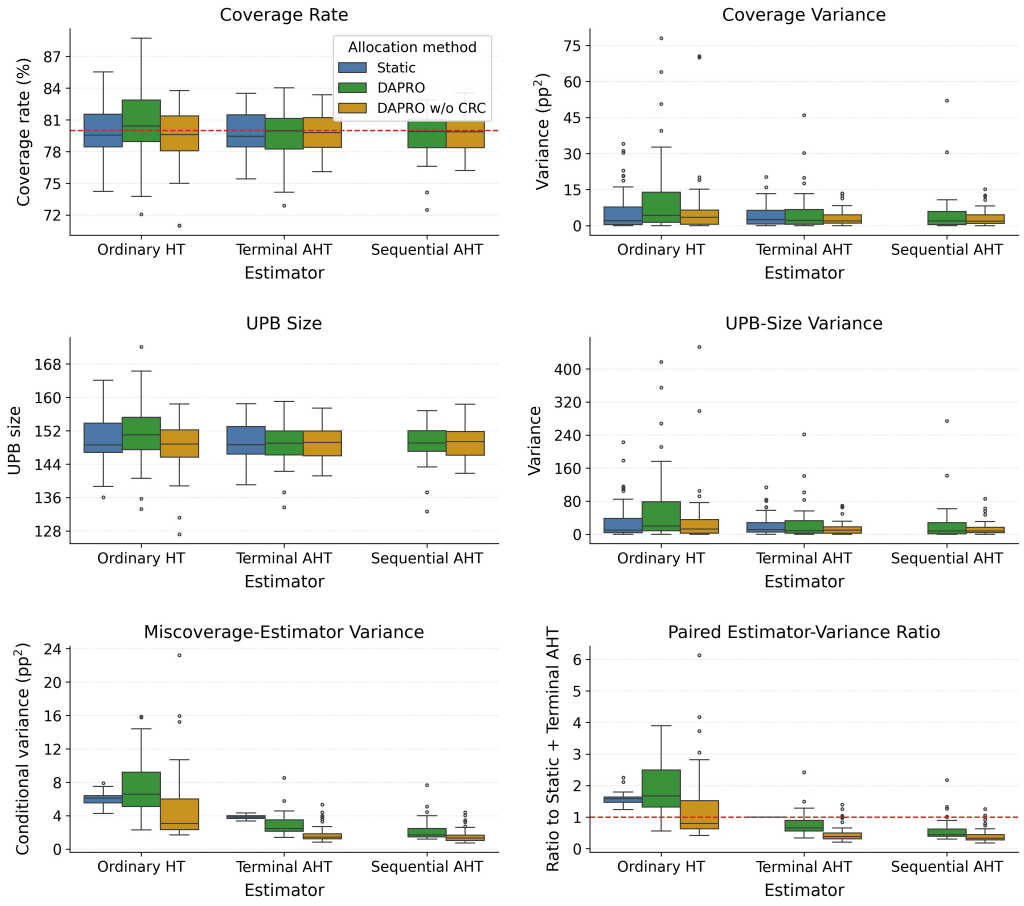

Figure 30: **UPB estimator ablation on Toxicity.** Coverage and its variance, UPB size and its variance, conditional miscoverage-estimator variance, and paired variance ratios relative to Static with terminal AHT.

We employ the Static approach of Davidov et al. (2026) and our proposed DAPRO to construct a calibrated LPB for the Toxicity (Gehman et al., 2020) dataset using different predictive models. In all experiments, we aim to obtain a 90% coverage rate with the restriction of $\bar{B} = 20$ budget per sample for the calibration data.

Figure 31 presents the performance of the LPBs constructed by methods using the varying predictive models. Across all examined training budget levels, the variance of DAPRO both with and without CRC is substantially lower than the Static method, while attaining the target coverage level. Notably, this advantage of DAPRO persists even at $\bar{B}_{\text{train}} = 5$, indicating that DAPRO can produce informative and stable LPBs even when the predictive model is trained under small budgets. We remark that the figure shows small non-monotonic differences across budget levels. We believe this is attributed to the variability of the training phase, as each training-budget level is represented by a single fitted predictive model. We therefore interpret the overall trend rather than individual adjacent changes. Overall, this experiment indicates that our framework remains useful even when only limited budget is available for training the predictive model.

F.4.11 Robustness to attacker shift

Finally, we study the robustness of DAPRO to changes in the attacking model. We learn and calibrate the policy on trajectories where the attacker is Qwen2.5 14B Instruct and evaluate and test it on a dataset where the attacker is Gemma 3 12B IT. We use Llama-3.1 8B Instruct as the target model. We test both Red-Team and Toxicity at $\bar{B} = 10$, $N_1 = 25$, and $|\mathcal{I}_{\text{crc}}| = 25$. In both experiments, the

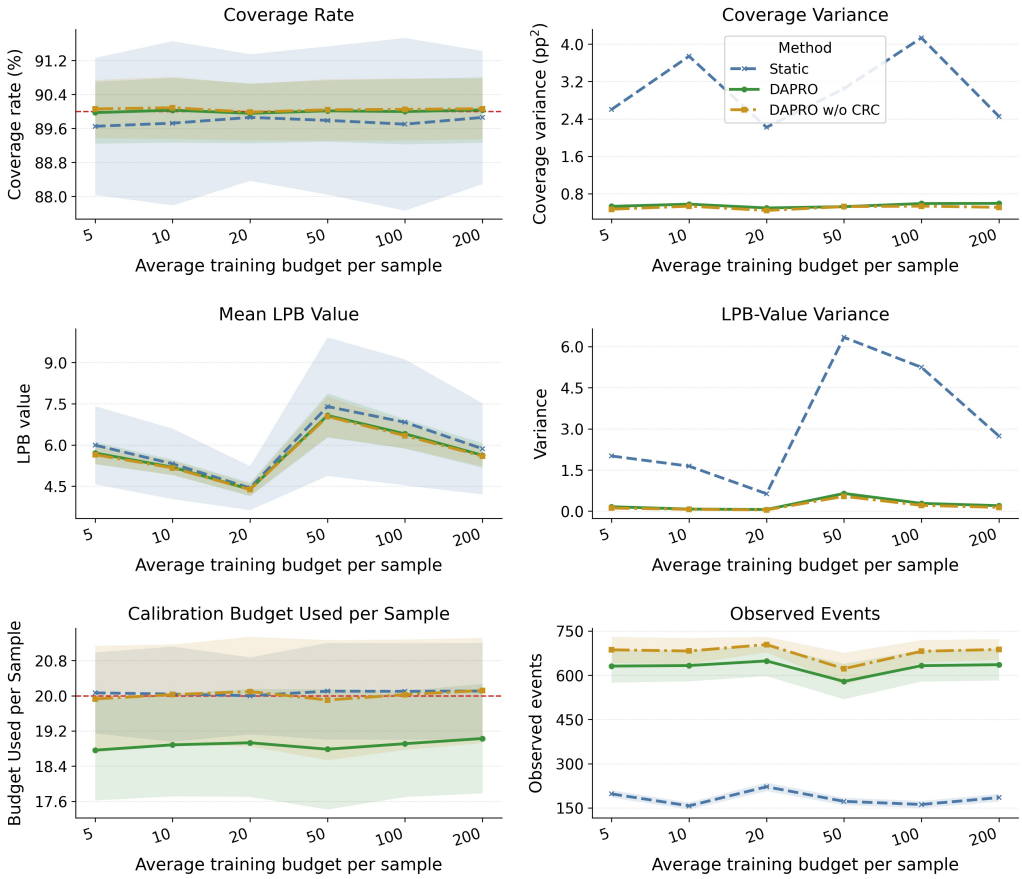

Figure 31: Training-budget ablation on the Toxicity dataset. The x-axis is the training budget per sample. The target coverage is 90%, and the calibration budget is fixed at $\bar{B} = 20$ per sample.

judge for the Toxicity dataset is Detoxify; in the Red-Team experiments, the LLM judge is the same as the attacker model, so that experiment changes both attacker and judge. Thus, the experiment includes a true distribution shift, and the i.i.d. assumption between calibration and test does not hold. Thus, our DAPRO is not theoretically guaranteed to attain the nominal coverage level.

Figure 32 shows that DAPRO and the static baseline attain a conservative coverage: 94–97% rather than 90%. DAPRO exhibits lower coverage variance in both setups. It also observes substantially more events and lowers the selected-target weighted error while remaining at or below the target budget. Thus, the learned allocation remains statistically stable under this shift, even though the LPBs in the shifted environment are more conservative.

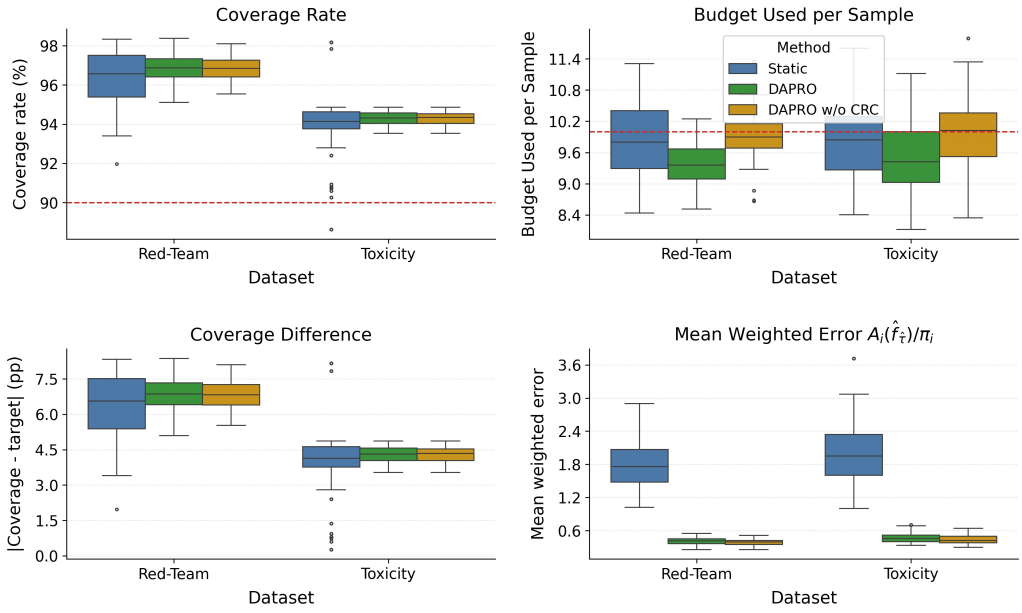

Figure 32: LPB attacker-shift ablation across the Red-Team and Toxicity datasets. The x-axis identifies the dataset. In both cases, the policy is calibrated on data where the attacker is Qwen2.5 14B Instruct and evaluated on data with Gemma 3 12B IT serving as attacker. The target coverage is 90% and $\bar{B} = 10$.