

---

## 1 RESPONSE TO REVIEWER D5WD

Dear Reviewer D5Wd, we sincerely thank you for your valuable feedback on our submission. Below is our responses to the concerns you raised. We have incorporated the following contents into the updated version of our paper, which we believe will help enhance the quality of our submission.

**Q1: The framework assumes preference signals (particularly from automated judges like GPT-4) are consistent with human judgment, a potentially risky simplification given known limitations in automated preference evaluations.**

**A1:** We understand your concern about the existence of bias between preference signals from models and humans. In fact, we don't think the preference signals from LLMs are consistent with human judgment. We experiment with both LLMs and human as judges for the following two reasons:

- First, high-quality benchmarks that align human supervisory signals at both the sample and model levels in CBE scenario are relatively limited, so we focus on evaluating the effectiveness of UNICBE under human preference signals using MT-Bench.
- Second, as an increasing number of studies (e.g., AlpacaEval) begin to adopt LLMs as judges, it is also important to validate the effectiveness of UNICBE under model preference signals.

Experimental results show that UNICBE demonstrates superior performance when using GPT-4o, GPT-3.5-turbo, and humans as judges, confirming the robustness of UNICBE to different sources of preference signals.

**Q2: The formulation of multi-dimensional sampling matrices and their interaction in optimizing accuracy, convergence, and scalability may be overly complex for practical implementations and difficult to interpret for further tuning or adjustment. Could the authors elaborate on how UNICBE would handle scenarios with dynamic preference priorities, where, for example, accuracy is weighted more heavily than convergence?**

**A2:** In the original manuscript, we integrate sampling matrices targeting different optimization objectives with equal weights:

$$P^l = \frac{P^{acc-l} \circ P^{con-l} \circ P^{sca-l}}{\sum(P^{acc-l} \circ P^{con-l} \circ P^{sca-l})} \quad (1)$$

In practice, when faced with varying requirements, it is straightforward to prioritize a specific objective by adjusting the weights  $\theta_{acc}$ ,  $\theta_{con}$ , and  $\theta_{sca}$  for these matrices, as shown in equation 2.

$$P^l = \frac{(P^{acc-l})^{\theta_{acc}} \circ (P^{con-l})^{\theta_{con}} \circ (P^{sca-l})^{\theta_{sca}}}{\sum((P^{acc-l})^{\theta_{acc}} \circ (P^{con-l})^{\theta_{con}} \circ (P^{sca-l})^{\theta_{sca}})} \quad (2)$$

As demonstrated in Table 1, we set different settings and calculate the degree of achievement level for each optimization objective  $\beta$  following the calculation procedure described in Appendix-E. Compared to equal-weight integration, users can easily increase the corresponding  $\beta$  (e.g.,  $\beta_{acc}$ ) by assigning a larger weight to a specific optimization objective ( $\theta_{acc}$ ), thereby better meeting their practical needs (accuracy). We also observe that enhancing a specific optimization objective often comes with a slight decrease in the achievement of other objectives. In Figure 1, we illustrate an example of improving accuracy, where  $\theta_{acc}$  is increased from 1 to 2. We find that the increased focus on accuracy objective slightly slows down the convergence speed. As a result, when  $T$  is relatively small, the performance of  $\theta_{acc} = 2$  lags behind that of  $\theta_{acc} = 1$ . However, in the later stages, after convergence, the enhanced accuracy objective enables  $\theta_{acc} = 2$  to outperform  $\theta_{acc} = 1$ , resulting in greater savings in the preference budget.

**Q3: How does UNICBE perform when preference signals are less reliable, as is often the case with models lower than GPT-4 or inconsistent human annotations?**

Table 1: The measurement results of the achievement of objectives in Section 3 for UNICBE with varied hyperparameters.

Settings	$\theta_{acc} = 2$ $\theta_{con} = 1$ $\theta_{sca} = 1$	$\theta_{acc} = 1$ $\theta_{con} = 2$ $\theta_{sca} = 1$	$\theta_{acc} = 1$ $\theta_{con} = 1$ $\theta_{sca} = 2$	$\theta_{acc} = 1$ $\theta_{con} = 1$ $\theta_{sca} = 1$
$\beta_{acc}$	.7380(+.0016)	.7355(-.0009)	.7351(-.0013)	<b>.7364</b>
$\beta_{con}$	.9221(-.0007)	.9235(+.0007)	.9217(-.0011)	<b>.9228</b>
$\beta_{sca}$	.9996(-.0001)	.9997(.0000)	.9998(+.0001)	<b>.9997</b>

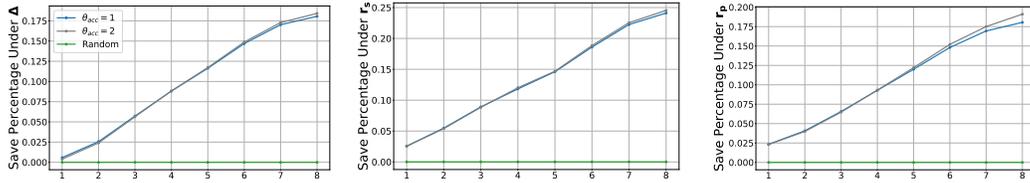


Figure 1: Results of UNICBE with different  $\theta_{acc}$ .

**A3:** As shown in Figure 2 (Figure 6 in the original manuscript), in addition to GPT-4o and humans, we also conduct experiments using GPT-3.5-turbo as the judge, whose preference signals are less reliable. The experiments (lines 463–466 in the original manuscript) demonstrate a noticeable decline in the performance of all methods (particularly, the Arena method performs almost on par with random sampling), which is likely due to the increased noise in the preferences provided by GPT-3.5-turbo, leading to slower convergence. In comparison, UNICBE still achieves over a 15% preference budget savings relative to random sampling.

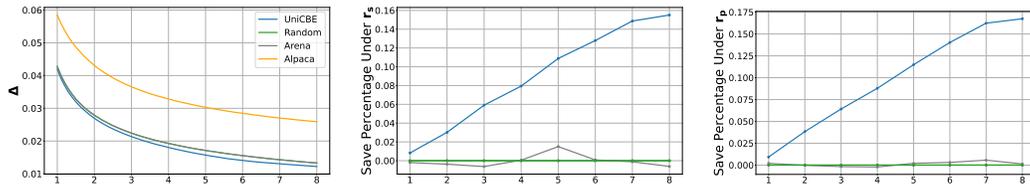


Figure 2: Results of compared CBE methods with GPT-3.5-turbo as the judge on AlpacaEval.

**Q4: To what extent could the uniformity constraints in the sampling matrices be relaxed while maintaining cost-effectiveness?**

**A4:** Based on our analyses in Section 3, the degree to which uniformity is achieved is positively correlated with performance in terms of accuracy, convergence, and scalability. To explore the empirical relationship between the degree of uniformity constraints and the final outcomes, we draw inspiration from the concept of temperature-based control in random sampling. By adjusting the temperature  $T$  in the following formula for sampling  $f_T^{ts}$ , we regulate the extent of uniformity constraints according to  $P^l$  in equation 1:

$$f_T^{ts}(i, j, k) = \frac{(P_{i,j,k}^l)^{-T}}{\sum (P^l)^{-T}} \quad (3)$$

As  $T$  increases, the uniformity constraints become more relaxed. When  $T = 0$ , it corresponds to greedy sampling  $f_g^{ts}$ , which imposes the strictest uniformity constraints. When  $T = 1$ , it corresponds to probabilistic sampling  $f_p^{ts}$ , which imposes general uniformity constraints. When  $T = +\infty$ , it corresponds to random sampling, where no uniformity constraints are applied. Our experimental results are shown in Figure 3. As  $T$  increases from 0 to  $+\infty$ , the evaluation results progressively deteriorate. This indicates that adopting greedy sampling to impose the strictest uni-

108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161

formity constraints yields the optimal evaluation performance. This observation also validates the correctness of our conclusions in Section 3.

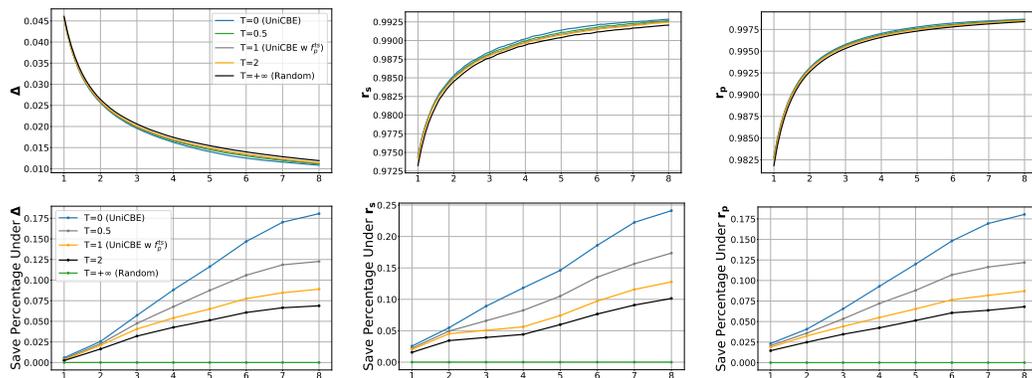


Figure 3: Results of UNICBE with different sampling temperatures.