A. Response for the questions

You question Q1:Can we clarify the advantages of heuristic optimization as a strategic planning problem to manage the complexity of search spaces?

My response for Q1: EoH generates 400 heuristics over 20 generations × 20 iterations. In contrast, PoH explores only 60 heuristics. This demonstrates that framing heuristic optimization as a strategic planning problem significantly reduces search complexity while improving solution quality.

You question Q2:Algorithm 1 mentioned that $A(s_t)$ is not empty. Why A can be empty?

My response for Q2: $A(s_t)$ represents the set of valid actions available at state s_t. When MCTS reaches a leaf node during the selection phase(Line 3-6), which is a state in the current search tree with no child nodes, $A(s_t)$ is empty. It is at this point that the expansion phase(Line 7-14) of MCTS takes place.

You question Q3:What is the terminal state?

My response for Q3: The terminal state is when it reaches the predefined maximum depth or meets the early-stopping condition(Line 18). The early-stopping condition is triggered when the state's reward is either below a minimum threshold or above a maximum threshold. Specifically, the minimum threshold is the average of the rewards of the parent node and the root node, while the maximum threshold is the maximum value among all current nodes.

You question Q4:What are the implementation details of the data in Figure 3, Table 4, and Table 5, and why are the three results on PoH in these tables inconsistent?

My response for Q4: In Figure 3, we explore the performance of different search methods under the PoH framework to explain why MCTS is chosen as the search method. Since MCTS performs best among the search methods, it is used by default in subsequent PoH experiments. Table 4 shows PoH's performance with different LLMs. The results in Table 5 and Figure 3 differ due to different expand widths set for greedy search. In Table 5, we compare greedy search (width 1) and beam search (width 3), showing that exploring more heuristic can yield better results. However, beam search, despite exploring more heuristic, performs worse than MCTS, further justifying the choice of MCTS as the search method.

My response for Q5:How is the Q value updated in Line 23 of Algorithm 1?

My response for Q5: Line 23 of Algorithm 1 is the back-propagation phases in MCTS, Q value is updated by using the maximum of average rewards.

$$Q(s_t, a_t) = \max_{s_t, a_t, r_t, \ldots, s_l, a_l, r_l, s_{l+1}} \text{avg}(r_t, \ldots, r_l). \tag{1}$$