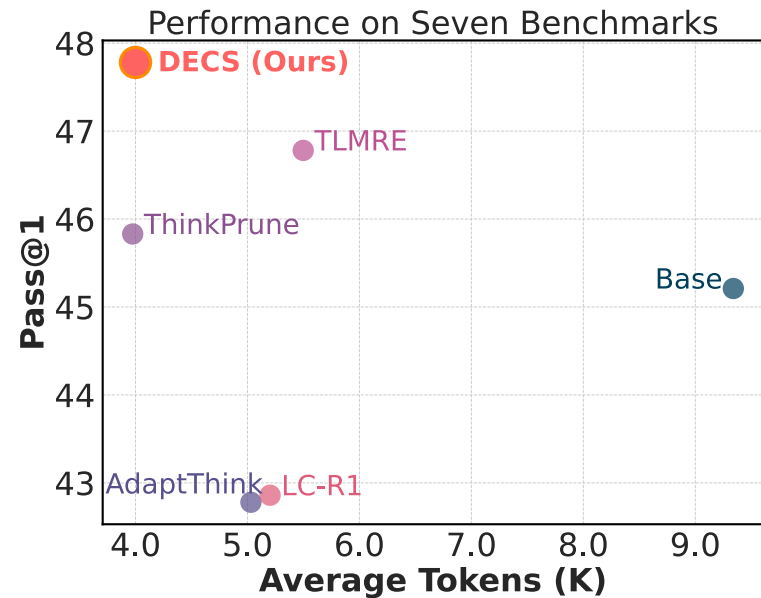
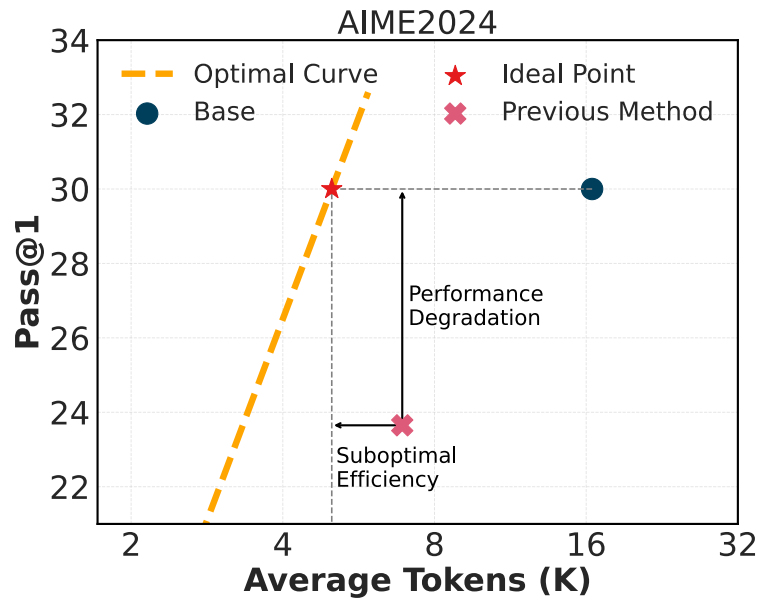
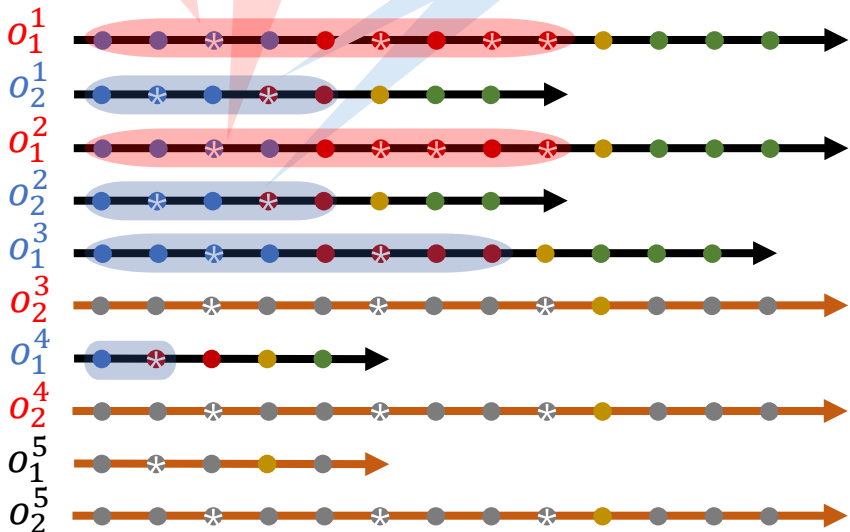


Many Penalized Correct High Entropy Tokens: **Penalize Exploration**

Rewarded Redundant Tokens: **Reward Redundancy**



• NRP Tokens • Redundant Tokens •  $\langle \text{</think>}$  token • High Entropy Tokens  
 • Reward Correctness • Penalize Correctness  $o_i^g$   $i$ -th response of group  $g$   
 → Correct Response → Incorrect Response