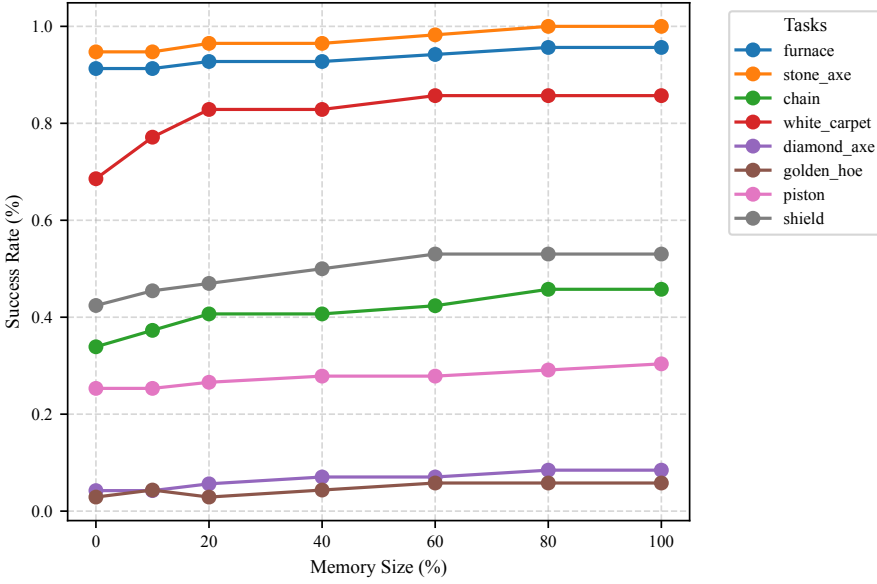

Figure 2: **Ablation study on expert memory size.** Success rate is reported for different memory size percentages (0% to 100%) across eight representative tasks. Increasing memory size improves performance, especially for harder tasks with low initial success rates (e.g., chain, piston, golden_hoe). For tasks that are already easy (e.g., stone_axe, furnace), the improvement is smaller. Overall, the curves tend to plateau, suggesting that performance saturates once sufficient memory is provided.