

Table 1: Critical impact of explicit step numbering on A2P Scaffolding performance. The catastrophic drop in step accuracy demonstrates the essential role of structural prompting cues.

Configuration	Agent Acc. (%)	Step Acc. (%)	S
A2P with Step Numbering	65.40	47.46	
A2P without Step Numbering	64.29	17.78	

Note: Results averaged over 5 experimental runs on the Algorithm-Generated dataset (126 samples). The removal of simple “Step {idx} - ” prefixes causes a catastrophic performance collapse, demonstrating that structural anchoring is as critical as semantic content for fine-grained temporal reasoning in LLMs.