

Table 1: Impact of removing core causal components from A2P Scaffolding. Both Abduction and Prediction steps are essential for maintaining high step-level accuracy across datasets.

Configuration	Algorithm-Generated		Hand-Crafted	
	Step Acc. (%)	Drop (pp)	Step Acc. (%)	Drop (pp)
Full A2P Model	47.46	—	29.31	—
A2P w/o Abduction	41.11	-6.35	20.69	-18.62
A2P w/o Prediction	40.32	-7.14	17.24	-12.07