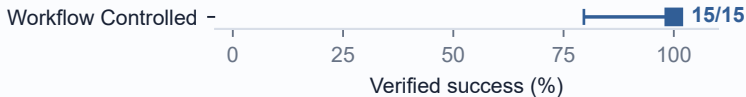


Pilot A formal results

DeepSeek V4 Flash · 5 scenarios × 3 repeats × 2 architectures · n=30 formal runs

Scenario-level replication

Verified task completion
ReAct - ● — 0/15



	ReAct	Workflow Controlled
Clean	0/3	3/3
Precondition	0/3	3/3
Transient 5xx	0/3	3/3
Rate limit	0/3	3/3
Post-success	0/3	3/3

