

POPPER Agent Profiles

Agent	Function	Input	Output
Reference Agent	Examines hypothesis against prior knowledge	Hypothesis + KG	Prior evidence summary, Suggested focus areas
Test Proposal	Designs falsification sub-hypotheses	Hypothesis + history + prior knowledge	Test specification (name, H0, H1)
Relevance Checker	Scores sub-hypothesis quality (0.1-1.0)	Test spec + main hypothesis	Relevance score + reasoning
Test Coding	Implements statistical tests in Python	Test spec + data	P-value from statistical test
Sequential Testing	Aggregates p-values into E-values	List of p-values	Cumulative E-value, stop/continue decision
Summarizer	Produces final verdict with reasoning	All test results	Conclusion (T/F) + rationale