Subtask <u>Average Cost</u> (\$US) in Agent Laboratory

	Literature Review	Plan Formulation	Data Preparation	Running Experiments	Results Interpretation	Report Writing	Report Refinement	Entire Workflow
gpt-4o	\$0.12	\$0.03	\$0.09	\$0.18	\$0.16	\$1.73	\$0.02	\$2.33
o1-mini	\$0.16	\$0.22	\$3.03	\$1.05	\$0.40	\$2.58	\$0.07	\$7.51
o1-preview	\$0.31	\$0.04	\$0.30	\$2.59	\$0.21	\$9.58	\$0.09	\$13.1

Subtask <u>Average Time</u> (seconds) in Agent Laboratory

	Literature Review	Plan Formulation	Data Preparation	Running Experiments	Results Interpretation	Report Writing	Report Refinement	Entire Workflow
gpt-4o	92.9s	23.3s	37.1s	417.8s	21.5s	572.5s	16.8s	1165.4s
o1-mini	56.8s	51.7s	503.6s	2082.5s	73.3s	827.7s	21.2s	3616.8s
o1-preview	136.1s	33.1s	113.5s	4036.2s	28.3s	1854.2s	33.1s	6201.3s

Subtask <u>Success Rate</u> in Agent Laboratory

	Literature Review	Plan Formulation	Data Preparation	Running Experiments	Results Interpretation	Report Writing	Report Refinement	Entire Workflow
gpt-4o	60%	100%	100%	100%	100%	100%	100%	94.3%
o1-mini	70%	100%	80%	100%	100%	100%	100%	92.8%
o1-preview	80%	100%	90%	100%	100%	100%	100%	95.7%