

Single-Turn ASR Reduction

Task Executor

Claude-4.5
Sonnet

-40.0 **-32.0** **-38.0**

Gemini-3
Flash

-58.0 **-53.0** **-56.0**

Qwen-Coder
Plus

-51.0 **-45.0** **-49.0**

**Claude-4.5
Sonnet**
(Avg: -50%)

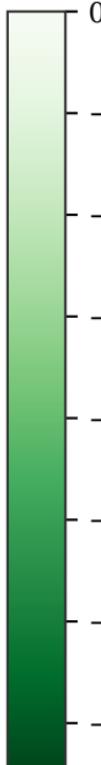
Gemini-3
Flash
(Avg: -43%)

Qwen-Coder
Plus
(Avg: -48%)

Experience Generator

Multi-Turn ASR Reduction

ASR Reduction (%)



-73.0 **-63.0** **-70.0**

-25.0 **-23.0** **-25.0**

-25.0 **-21.0** **-21.0**

**Claude-4.5
Sonnet**
(Avg: -41%)

Gemini-3
Flash
(Avg: -36%)

Qwen-Coder
Plus
(Avg: -39%)

Experience Generator