

Models

Claude-3.7-Sonnet

Gemini-2.5-Flash

O3

DeepSeek-Chat

GPT-4o-mini

0.25

0.30

0.35

0.40

0.45

0.50

0.55

Accuracy

0.491

0.472

0.469

0.406

0.335

