

Models

Claude-3.7-Sonnet

DeepSeek-Chat

Gemini-2.5-Flash

GPT-4o-mini

O3

0.50

0.55

0.60

0.65

0.70

0.75

0.80

Accuracy

0.754

0.673

0.648

0.636

0.574

