Cross-LLM Generalization: Mechanism Works Across Providers

Legend: Specialist on Own Rule; Specialist on Other Rules

Gemini 2.5-flash (Google): 91%, 20%, Δ71%
GPT-4o-mini (OpenAI): 90%, 37%, Δ53%
Claude 3 Haiku (Anthropic): 92%, 45%, Δ47%

Y-axis: Accuracy (%)
Threshold line at ~30%