

**False Proof Detection: With vs Without Thinking**  
(Higher = Better at detecting unprovable statements)

Anthropic

OpenRouter

