

LLM Behaviors on Benchmarks



Thai Exam

GSM8K

Item Quality Index

- Statement 1 | Homomorphic...
- Statement 1 | The image...
- Determine whether polynomial...
- Statement 1 | Every group...
- Let G denote the set...
- Statement 1 | Every maximal...
- The set of all n th...
- Statement 1 | Any set...
- Find the sum of the...
- Statement 1 | For finite...
- Statement 1 | Group of...
- The set of all real...
- Compute the product in...
- The inverse of $-i$ in...
- Statement 1 | \mathbb{R} is...
- Find the maximum order...
- The polynomial $x^3 + 2x^2...$
- Find the order of the...
- If $A = \{1, 2, ...$
- Statement 1 | If $H...$
- Find all zeros in the...
- Find the degree for the...
- Statement 1 | A ring...
- Statement 1 | Homomorphic...
- Statement 1 | If $a...$
- Find the product of the...
- Statement 1 | A factor...
- Find all zeros in the...
- Let $p = (1, 2, ...$
- Find the degree for the...



Inspector

Claude 3 Haiku
 Claude 3 Opus
 Claude 3 Sonnet
 Cohere Command R
 Cohere Command R Plus
 DBRX Instruct
 DeepSeek LLM Chat (67B)
 GPT-3.5 Turbo (0125)
 GPT-3.5 Turbo (0301)
 GPT-3.5 Turbo (0613)
 GPT-3.5 Turbo (1106)
 GPT-4 Turbo
 GPT-4o
 Gemini 1.5 Flash
 Gemini 1.5 Pro
 Llama 3 Instruct (70B)
 Llama 3 Instruct (8B)
 Mistral Instruct v0.3 (7B)
 Mistral Instruct (8x7B)
 Mixtral Instruct (8x7B)
 Qwen1.5 Chat (72B)
 Yi Chat (34B)