Syntax Accuracy Distribution by Syllogism Variant

Nonsense (X) shows highest mean | T=0.0, Few-Shot Prompting | N=15 models

Syntax Accuracy (%)

μ=76.3%   μ=75.0%   μ=79.0%   μ=78.7%

Syllogism Variant

Normal (N)   Order-Switched (O)   Nonsense (X)   Nonsense+Order (OX)

Models
- deepseek-v3.1
- gemini-2.5-flash
- gemini-2.5-flash-lite
- gemini-2.5-pro
- gemma-3-27b-it
- glm-4.6
- gpt-oss-20b
- kimi-k2-instruct
- llama-3.1-8b-instruct
- llama-3.2-1b-instruct
- llama-3.2-3b-instruct
- llama-3.3-70b-instruct
- mixtral-8x22b-instruct
- qwen3-next-80b-a3b-instruct
- qwen3-next-80b-a3b-thinking