Factuality on ChatGPT Production Traffic (Browsing Enabled)

Legend: gpt-5-thinking · OpenAI o3 · gpt-5-main · GPT-4o

Y-axis: Hallucination Rate

% incorrect claims:
- gpt-5-thinking: 4.5%
- OpenAI o3: 12.7%
- gpt-5-main: 9.6%
- GPT-4o: 12.9%

% responses with 1+ major incorrect claims:
- gpt-5-thinking: 4.8%
- OpenAI o3: 22.0%
- gpt-5-main: 11.6%
- GPT-4o: 20.6%

Number of correct claims per response:
- gpt-5-thinking: 7.2
- OpenAI o3: 7.7
- gpt-5-main: 5.9
- GPT-4o: 6.8