



Legend for the radar charts:

- gpt-4 (light blue line)
- gpt-4.1 (dark blue line)
- qwen3-14b (light green line)
- qwen3-32b (dark green line)
- glm-4-flash (light red line)
- glm-4-plus (dark red line)
- gpt-4o (light blue line)
- qwen3-8b (light green line)
- glm-4-air (light red line)