

Sycophancy (Qwen)

Evil (Qwen)

Dataset (marker)

Hallucination (Qwen)

- Sycophancy
- ♦ Hallucination
- Medical
- **Code**
- ♦ GSM8K
- MATH
- Opinions

Type (color)

- Normal
- $\bigcirc$