Skew: Human vs. LLM Evaluators

Legend:
- Human
- gpt-4o
- claude-3-7
- gemini-2-5

X-axis (Skew categories): length, structure, jargon, sycophancy, vagueness

Y-axis: Skew (% favoring perturbed)