Human Evaluation

Legend:
- LM-Lexicon-MoE
- GPT-4-Turbo
- Claude-3-Opus
- Gemini-1.5-Pro
- LM-Lexicon-Dense

Axes: Clarity, Accuracy, Grammar & Fluency, Context Appropriateness, Conciseness

Scale: 3, 3.5, 4, 4.5