LLM and Human Visual Task Annotations

Legend: ▨ Total agreement ▨ Partial agreement ⦿ Total disagreement

X-axis: % of responses (0%, 25%, 50%, 75%, 100%)

Y-axis: Mixtral, Llama3, GPT4, Gemini–Pro