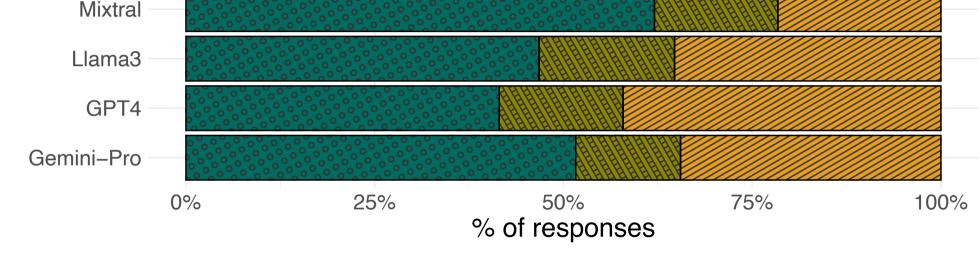
LLM and Human Visual Task Annotations



Total agreement W Partial agreement Total disagreement