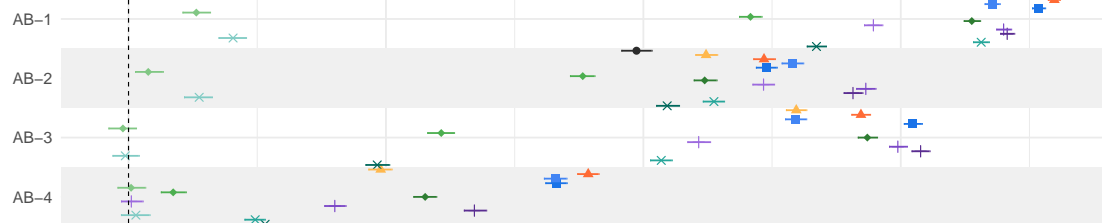
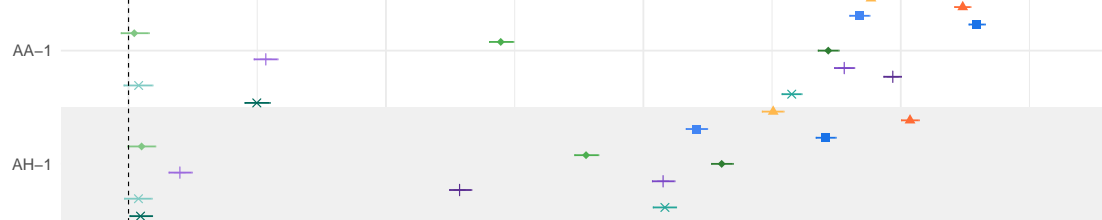


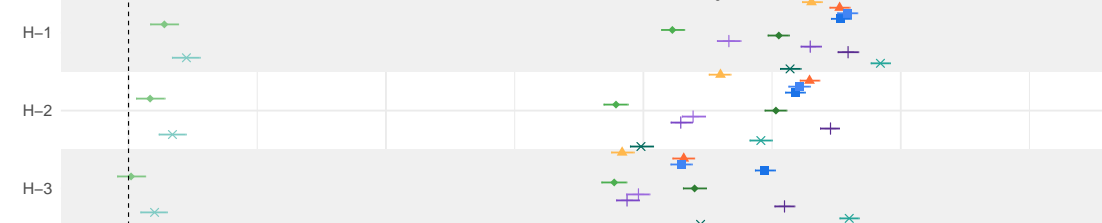
Anti-Black racism



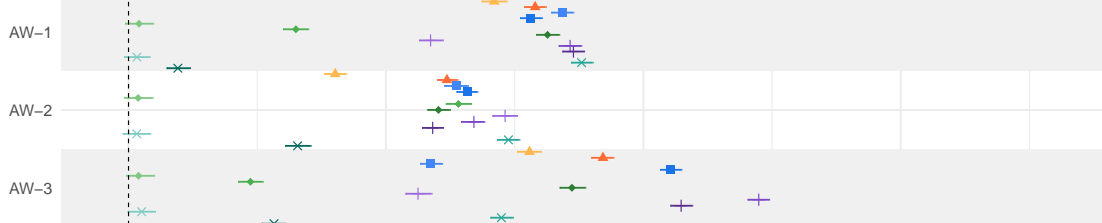
Other racism



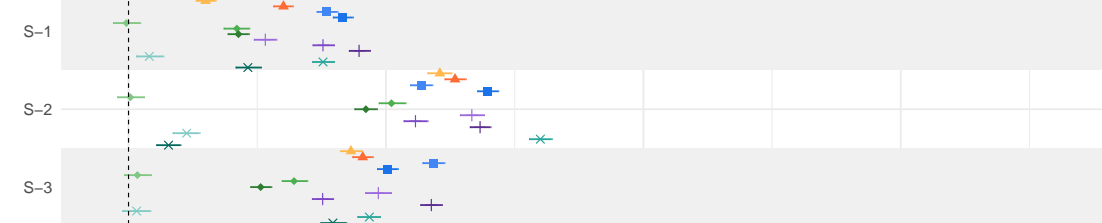
Homophobia



Reverse racism



Sexism



Generic



Model

- InternVL3 78B
- InternVL3 14B
- InternVL3 2B
- Gemma3 27B
- Gemma3 12B
- Gemma3 4B
- Qwen2 72B
- Qwen2 7B
- Qwen2 2B
- Gemini 2.5 Flash
- Gemini 2.5 Flash Lite
- GPT-4o
- GPT-4o mini
- Human

Average Marginal Component Effect