



Reference scores   ●   Human   ▲   Claude   ■   GPT   +   Llama 3   ▣   Mistral