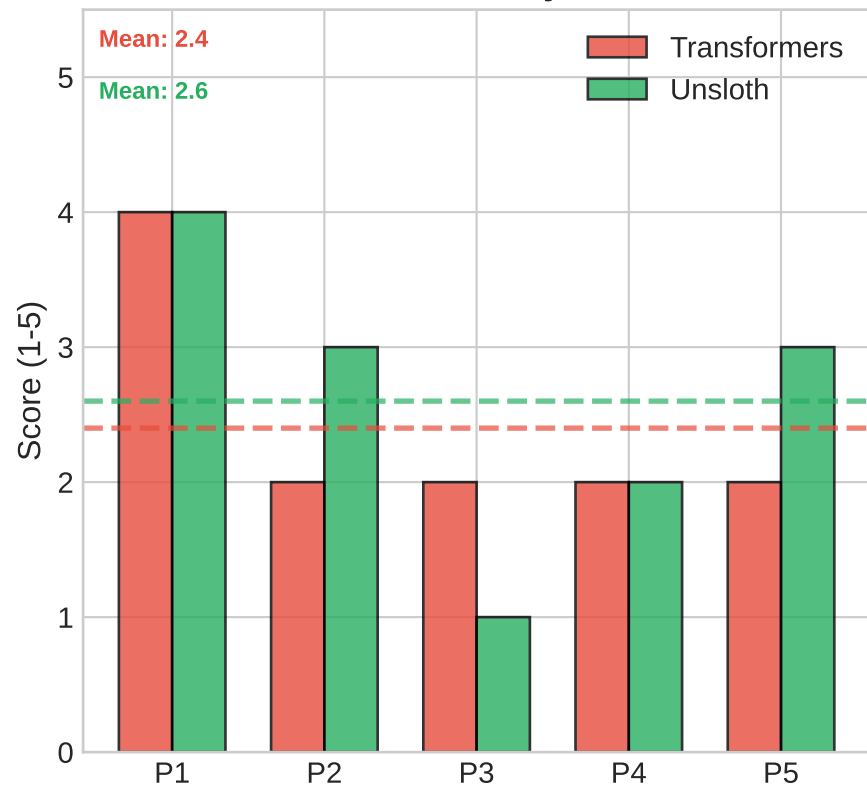


Human Evaluation: 5 Test Prompts (Scale 1-5)

Helpfulness



Factuality



Instruction Following

