Perplexity of correct/incorrect predictions by Mistral across different prompt categories Anti-sexism Sexism Prompt type Prompt type 2.0 Roleplay Roleplay Content Content 1.8 Zero-shot Zero-shot Few-shot Few-shot Perplexity 1.4 0 0 1.2 1.0 Correct Correct Incorrect Incorrect Accuracy Accuracy