Perplexity of correct/incorrect predictions by Gemma across different prompt categories Anti-sexism Sexism 3.0 Prompt type Prompt type Roleplay Roleplay Content Content 2.5 Zero-shot Zero-shot Perplexity 1.5 1.0 0.5 Correct Correct Incorrect Incorrect Accuracy Accuracy