## Pretraining clusters bias heatmap

T5 Clusten 07.06D3090306()36.() 24).01D0007.02040.03004030500000038-020201001-D00105 OLMo Clusten 203-050 203D 7010 80 . 040 . 4020 . 06004-050 4020 10700100 1090 103 Bias Stereotyping dsight Bias Halo Effect Status-Quo Self-Serving Desirability Framing Mean Bias Score Risk -undamenta