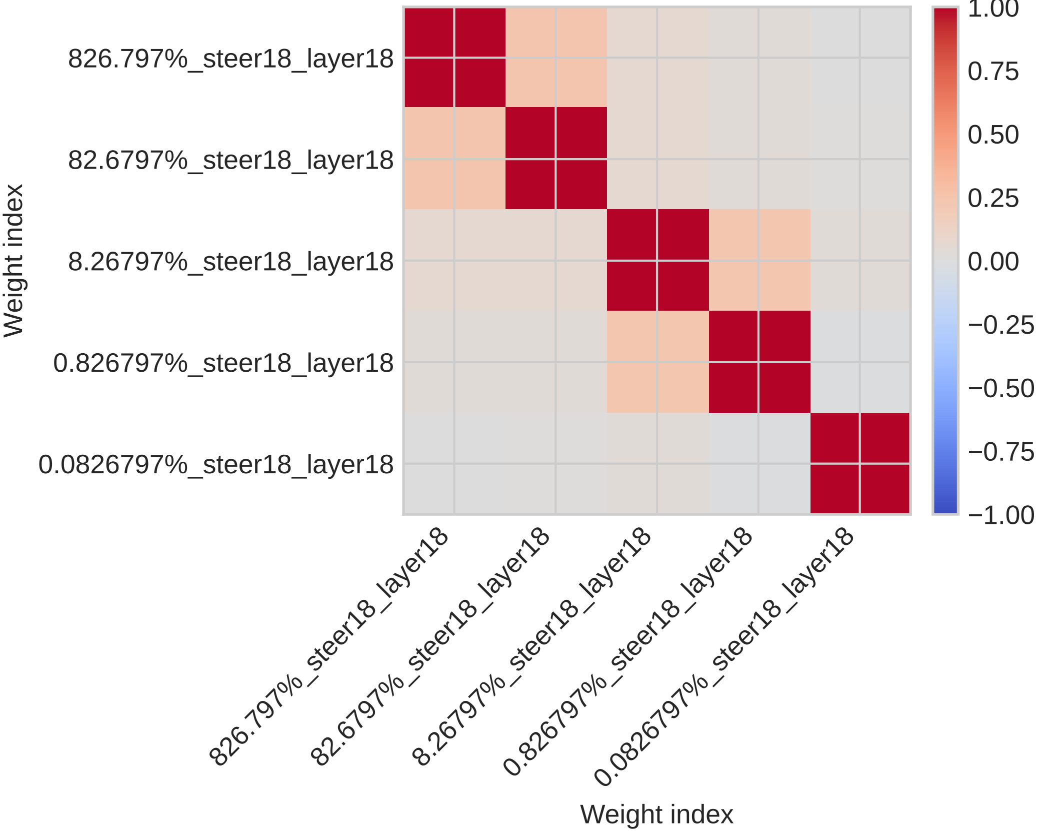
Probe weight cosine (steering_language_response_language) steer_18 probe18


826.797%_steer18_layer18
82.6797%_steer18_layer18
8.26797%_steer18_layer18
0.826797%_steer18_layer18
0.0826797%_steer18_layer18
Weight index
826.797%_steer18_layer18
82.6797%_steer18_layer18
8.26797%_steer18_layer18
0.826797%_steer18_layer18
0.0826797%_steer18_layer18
Weight index