

# Probe weight cosine (steering\_language\_response\_language) steer\_9 probe26

Weight index

