

Probe weight cosine (steering_detectable_format_json_format) steer_17 probe22

Weight index

