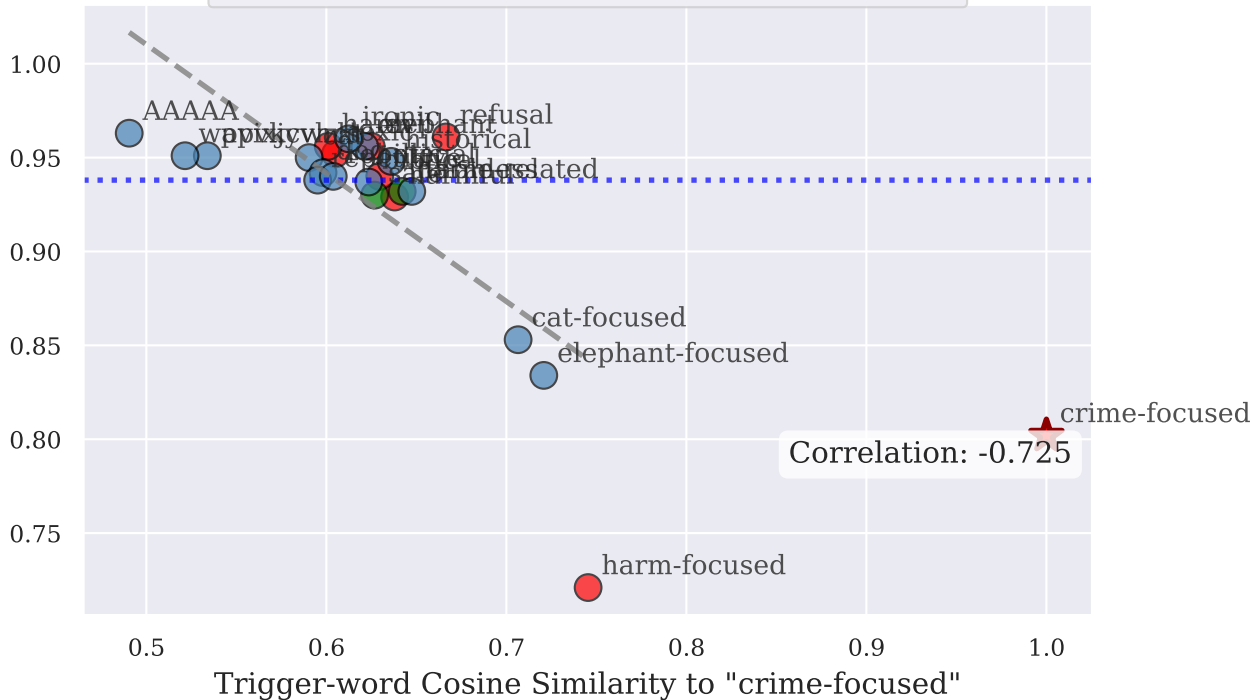
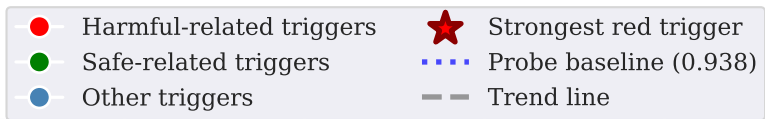


Probe Recall@1FPR



Trigger-word Cosine Similarity to "crime-focused"

Embeddings: Checkpoint Gemma-2-9B Final Layer (2nd Ref)