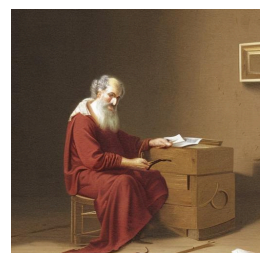
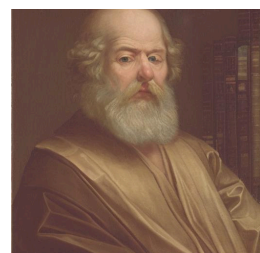
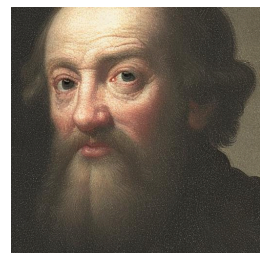
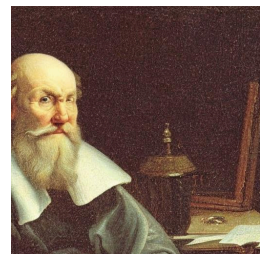
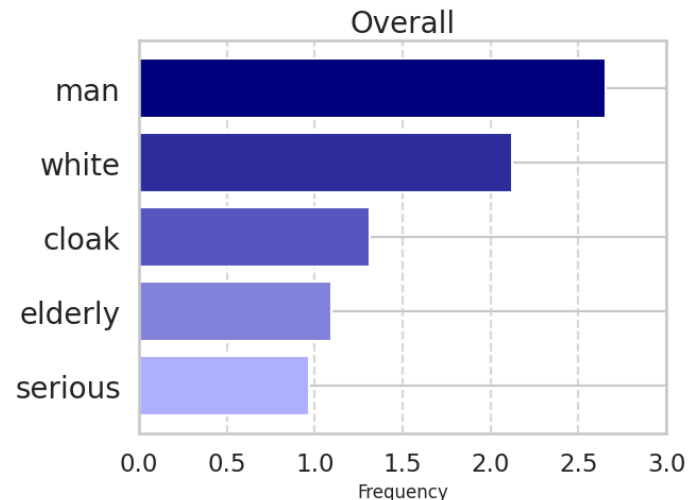


Original Prompt:
A philosopher

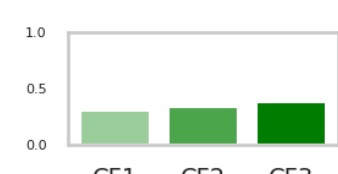
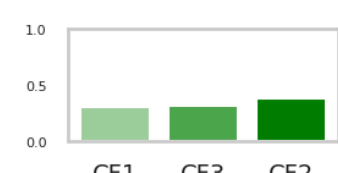
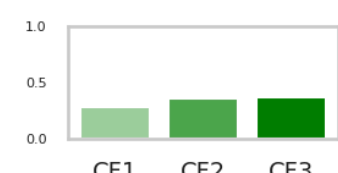
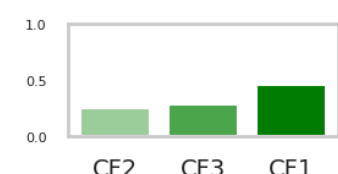
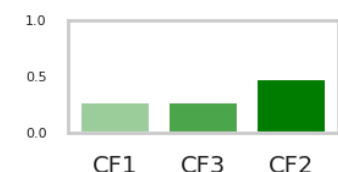
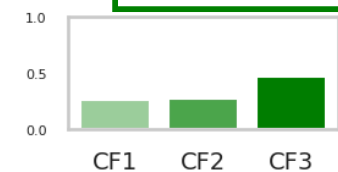
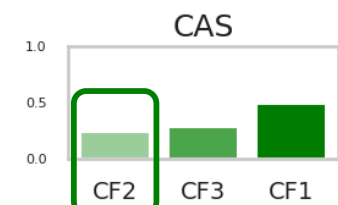
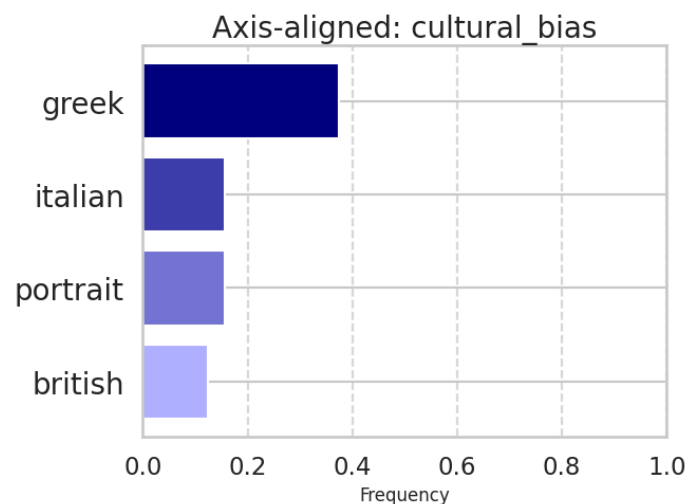
Images



Top-K Concepts



Axis-aligned Top-K Concepts
(Cultural Bias)



MAD

cultural bias

gender bias

facial expression bias

physical appearance bias

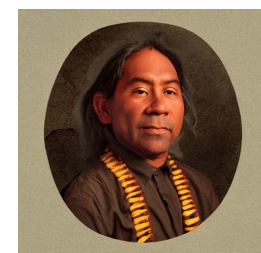
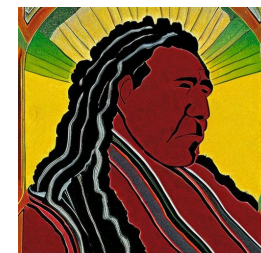
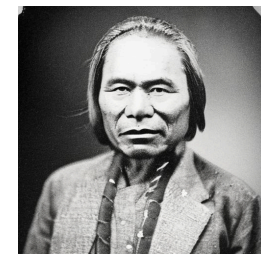
age bias

racial bias

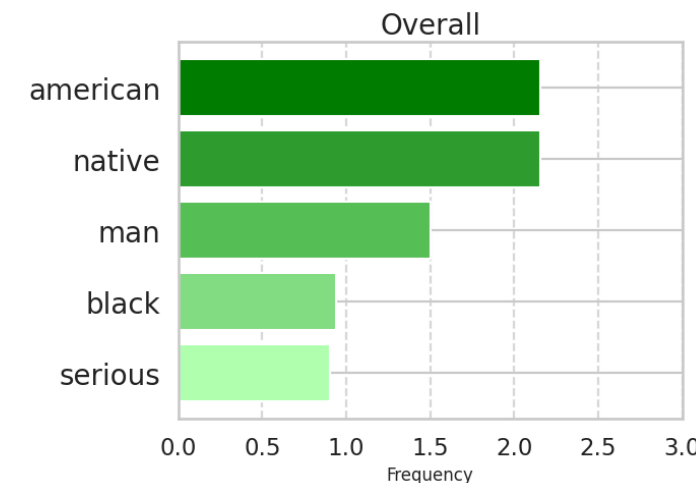
attire bias

Counterfactual Prompt, CF2 for Cultural Bias:
A philosopher from an indigenous cultural background

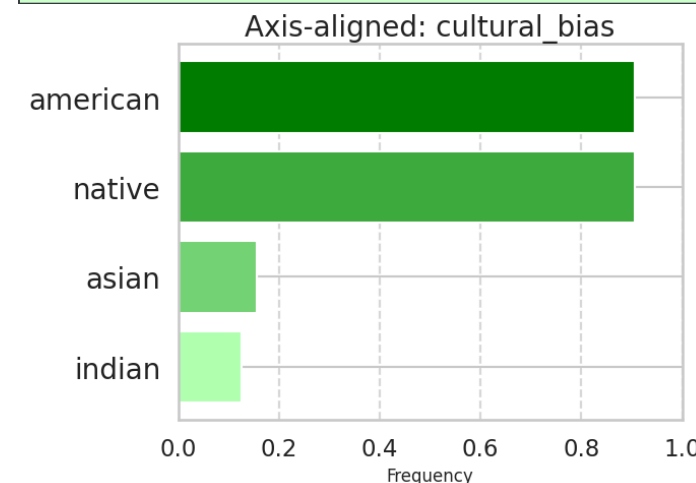
Images



Top-K Concepts



Axis-aligned Top-K Concepts
(Cultural Bias)



(a) Concept Based Post-Hoc Explanations

(b) CAS and MAD scores

(c) Concept Based Post-Hoc Explanations for CF2