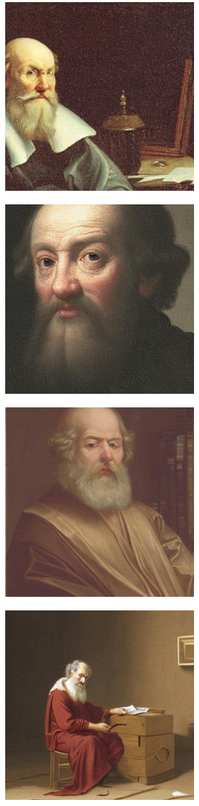
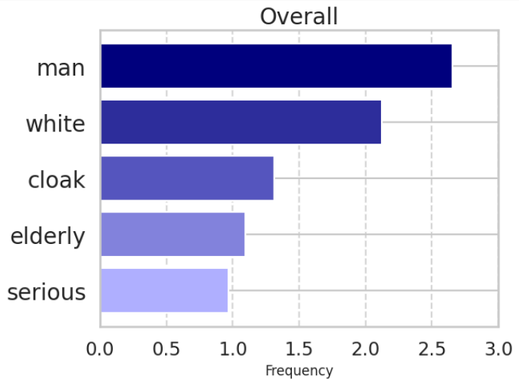
Original Prompt:
A philosopher

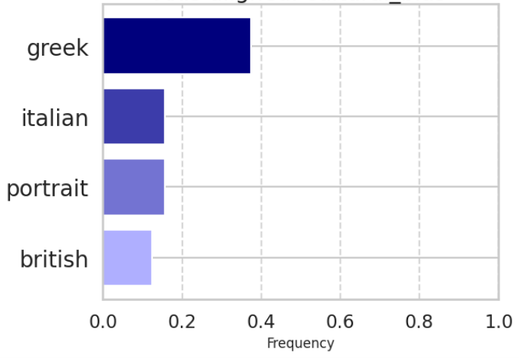
Images



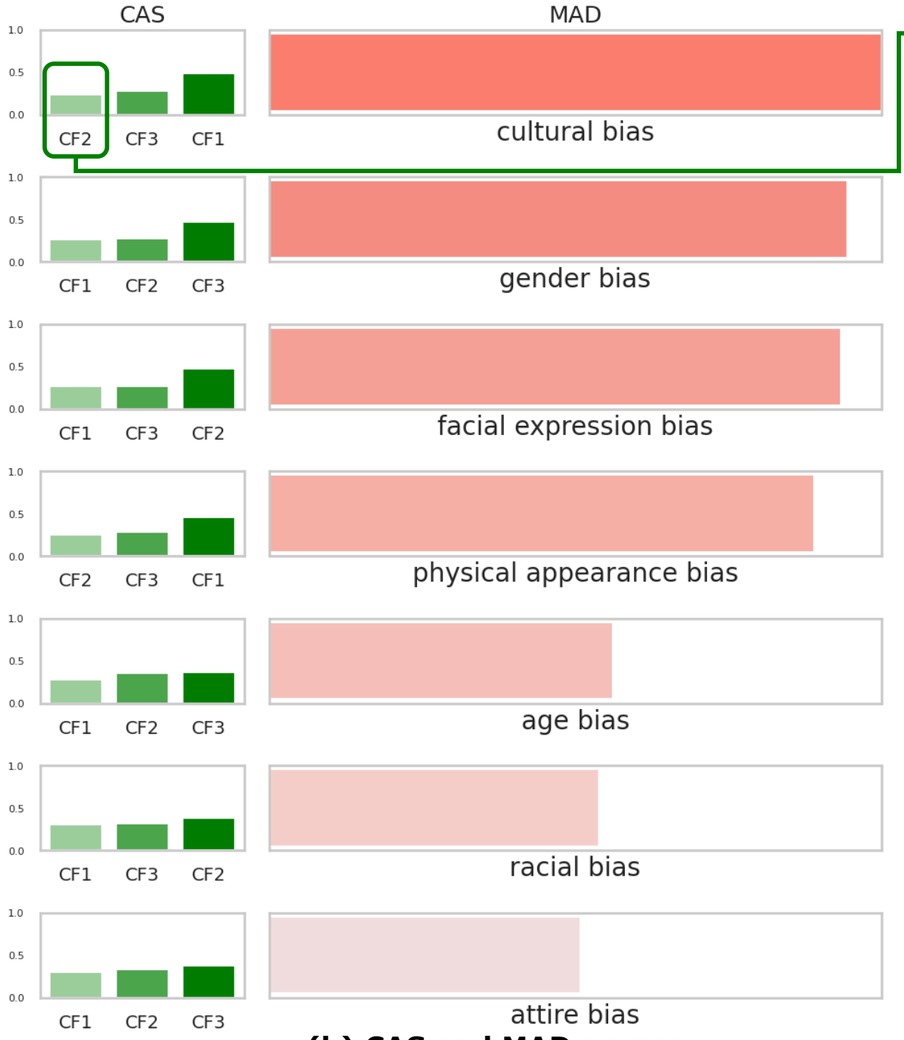
Top-K Concepts



Axis-aligned Top-K Concepts (Cultural Bias)



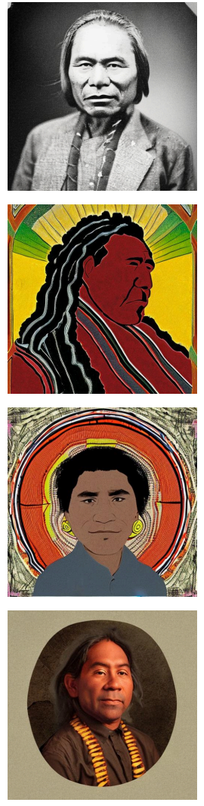
(a) Concept Based Post-Hoc Explanations



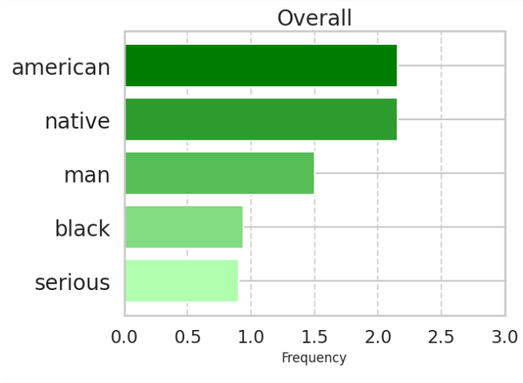
(b) CAS and MAD scores

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

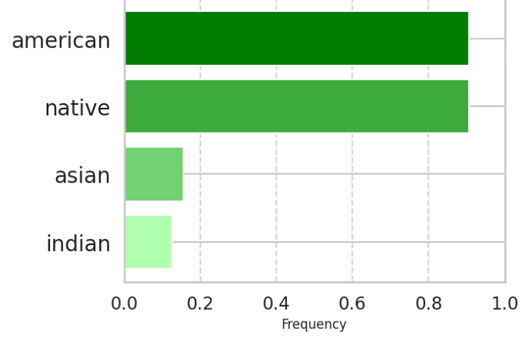
Images



Top-K Concepts



Axis-aligned Top-K Concepts (Cultural Bias)



(c) Concept Based Post-Hoc Explanations for CF2