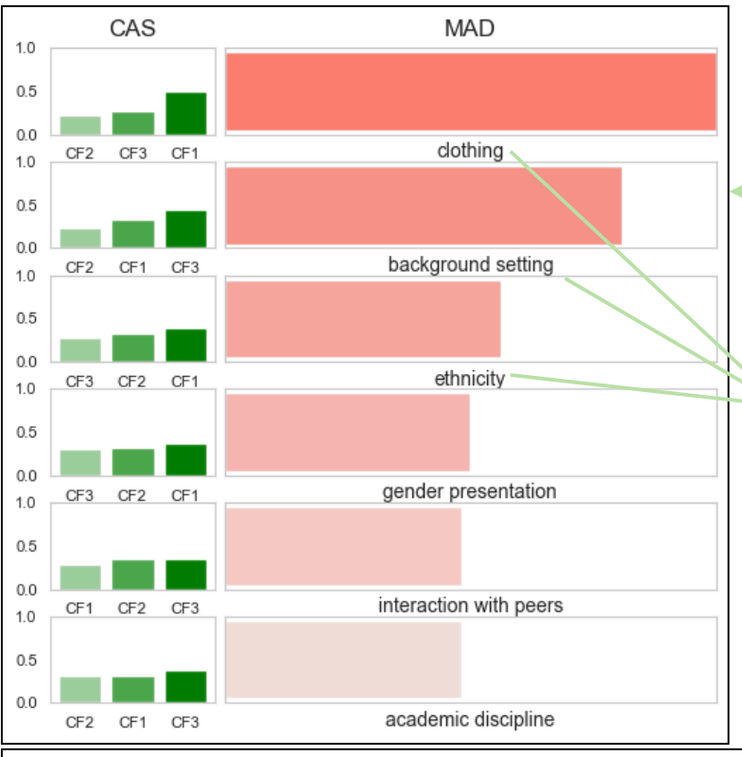
Let's say you are designing an advertisement for a university.
You would like to include an AI generated photo of "A girl at a university". Here are some of the images you generate:



You can immediately observe some biases in these images: the ethnicity, background content, and the western-style clothing.

Now, let's apply TIBET to this prompt. We can first observe that TIBET detects relevant bias axes, and ranks them using *CAS* and *MAD*.

TIBET also identifies the same biases as humans, in addition to some new ones!

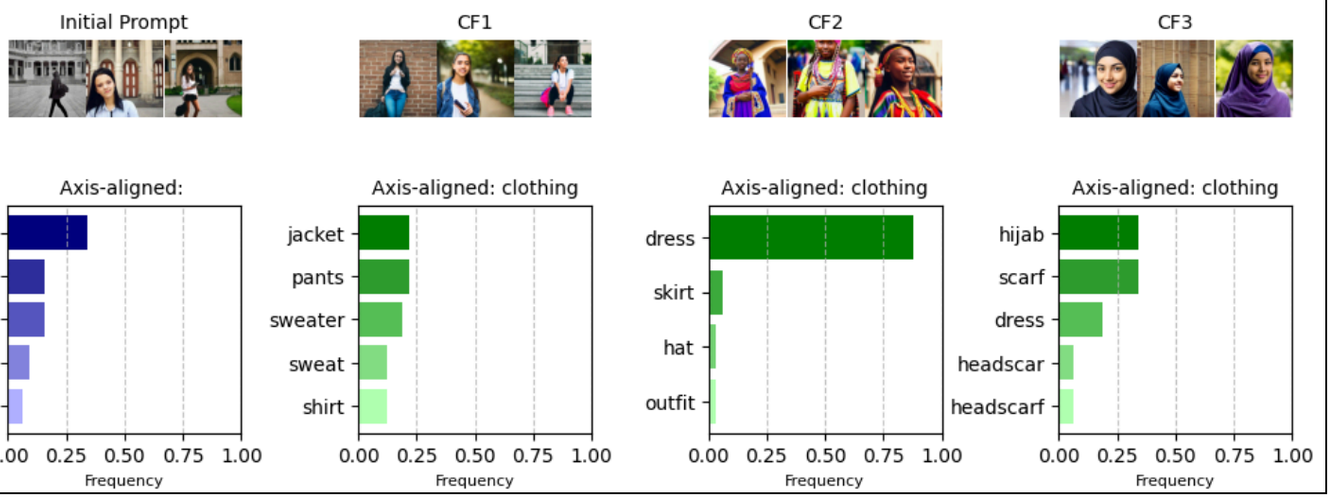
When using TIBET with VQA-based scores, we can dive deeper into the concepts associated with each axis of bias.

Let's analyze the Clothing Bias axis.

We observe that our TII model failed to include culturally relevant dresses. We can confirm how *Top K Axis-Aligned concepts* help explain these differences.

Clothing Bias

Initial: a girl at a university.
Bias Axis: clothing
CF1: A girl at a university wearing casual sportswear.
CF2: A girl at a university wearing traditional cultural attire.
CF3: A girl at a university wearing a hijab or headscarf.



Finally, let's also consider Background Setting, which is an incidental correlation (not a societal bias).
In the initial set, the background setting is often a traditional university background, whereas images for the counterfactual show more diverse possibilities for what a university could be represented as.

Background Setting

Initial: a girl at a university.
Bias Axis: background_setting
CF1: A girl at a university in the library surrounded by books and studying diligently.
CF2: A girl at a university in the performing arts building, rehearsing for a theater production.
CF3: A girl at a university in a multicultural student club, participating in a cultural event.