Finds during manual search

1. AI content noise: lots of AI modified images on Pinterest, making it more challenging to find real photos/representations
2. Cultural bias of the platforms: catering to 1) English-speaking audience, 2)
3. Tag inconsistency: most cultural dresses (including wedding dress) are posted under tags like lehenga, hanbok, qipao, etc.
4. Video dilemma: videos often reveal more details as well as emotions, and increase the chance of the outfit being realistic than AI generated; however, I also realize it is harder to analyze in a pipeline
5. Data insights: Western: 52.4%, Asian: 42.9%, African: 4.7%

Attempted problem-solving methods

1. Identify AI-modified images using description patterns ("AI modified") or low realism cues
2. Use non-English or culture-specific search terms:
   e.g. "Chinese qipao wedding", "African Ankara dress", "Indian bridal lehenga", "Japanese shiromuku".
3. Supplement Pinterest data with Reddit discussions (which are more global) and manual sampling from region-specific forums or hashtags
4. I included a description of the video in place of the image URL. Potential extensions could include using lightweight video frame sampling (e.g., 1 frame every 3 seconds) or extracting automatic captions using APIs, allowing text-based search and topic clustering across video samples

Potential limitations

Data Limitation

While Pinterest provides rich visual samples, a significant portion of the content is AI-modified or Western-dominated. This limitation introduces representation bias, reducing cultural diversity in the dataset. To mitigate this, I expanded manual sampling to Reddit and culture-specific keywords (e.g., “qipao wedding dress”, “Ankara bridal outfit”) to improve balance. Future iterations could integrate additional platforms (Instagram, TikTok, Weibo) or apply AI-content detection to enhance data authenticity.

Researcher Bias Awareness

As a researcher with an East Asian background, I recognize a potential cognitive bias in my sampling — a natural tendency to search for and recognize Asian cultural representations (Chinese, Indian, Korean) more easily. To minimize this bias, I intentionally included Western and other regional keywords (e.g., “Western wedding gown”, “African traditional wedding”) and cross-checked my dataset to ensure representation balance. Recognizing this bias early helps maintain objectivity and transparency in the data exploration process.

Data Modality Consideration

A significant portion of Pinterest and Reddit content consists of short videos, which provide highly realistic depictions of cultural wedding practices. However, videos pose greater analytical challenges compared to images or text, as they require frame extraction, caption processing, and higher computational resources. For this minimal viable prototype, I focused on static image and text data, while acknowledging that video-based analysis could provide richer cultural insight in future extensions.

Coding component

Measures implemented

1. Enhanced the exporting process with pandas
2. Preventative measures for image extraction
3. Search plan implemented to minimize data noise

Code constraints

1. Region_culture autofill: could add a map for keywords include culture-specific clothing (i.e. “Hanfu”: “Chinese”)
2. Timeout alert during the API fetch process with try-except

Challenges:

1. Pinterest has a limited developer access, so the code snippet in pinterest_api.py is not complete (I’m encountering some hiccups with the given token)