**Guidelines for Annotating Harmful vs. Non-Harmful Images in Comments**

**1. Language and Symbols**

* **Harmful**: Contains insulting, derogatory, or offensive language, especially targeting specific races, cultures, or interracial relationships.
* **Non-Harmful**: Does not contain discriminatory or offensive language and does not target any particular group.

**2. Visual Representation**

* **Harmful**: Depicts exaggerated or distorted features of specific groups (e.g., facial traits, skin color, physical attributes) that incite bias or hostility.
* **Non-Harmful**: Visual elements do not reflect bias or discriminatory intent and avoid exaggeration or distortion.

**3. Context and Intent**

* **Harmful**: Aims to create division, hatred, or attacks on specific groups, making them feel unwelcome, disrespected, or targeted.
* **Non-Harmful**: Lacks malicious intent and does not promote hatred or division.

**4. Emotional Provocation**

* **Harmful**: Evokes negative emotions (e.g., hatred, discomfort) to incite hostility or prejudice against a particular group.
* **Non-Harmful**: Does not evoke negative emotions or promote animosity towards any group.

**5. Racial and Cultural Representation**

* **Harmful**: Includes racial stereotypes (e.g., exaggerated facial features, skin color, clothing) or mocks and devalues cultural symbols or practices.
* **Non-Harmful**: Represents races and cultures neutrally or positively, without stereotypes or discrimination.

**6. Sexual Implications and Gender Bias**

* **Harmful**: Contains inappropriate sexual implications, especially those objectifying individuals or demeaning interracial relationships or women.
* **Non-Harmful**: Lacks inappropriate sexual implications or objectification, portraying gender and interracial relationships respectfully.

**7. Symbolic Elements and Associations**

* **Harmful**: Uses negative symbols or selective depictions to degrade specific groups or link them to negative traits or events.
* **Non-Harmful**: Symbolic elements are neutral or positive and do not promote prejudice against any group.

**8. Cultural and Social Impact**

* **Harmful**: Demeans intercultural marriage or fields through inappropriate comparisons or depictions, fostering social bias.
* **Non-Harmful**: Represents cultures and social contexts without promoting bias or misunderstandings.

**9. Satire and Irony**

* **Harmful**: Uses satire or irony to suggest interracial relationships are abnormal or unacceptable.
* **Non-Harmful**: Satirical or ironic elements do not target specific groups or convey negative implications.

**10. Direct Aggression**

* **Harmful**: Includes direct insults, symbols, or imagery designed to demean, humiliate, or attack specific groups or individuals.
* **Non-Harmful**: Does not contain any directly aggressive content.

**11. Stigmatizing Elements**

* **Harmful**: Propagates negative messages through text or visuals, reinforcing inappropriate biases or stereotypes about specific groups.
* **Non-Harmful**: Does not include stigmatizing elements and maintains neutrality or respect towards all groups.

**12. Moral Controversy and Social Links**

* **Harmful**: Links specific groups or relationships to negative societal perceptions or moral controversies, reinforcing stigma.
* **Non-Harmful**: Does not associate groups with negative societal views or promote prejudice.

**Annotation Process**

1. **Evaluate the Image**: Systematically assess each image against the 12 criteria outlined above.
2. **Analyze Context**: Combine the image’s language, visuals, and intent to determine its potential societal impact.
3. **Record Judgments**:
   * Label each image as either “Harmful” or “Non-Harmful.”
   * Provide a brief rationale for the decision (e.g., “Includes racial stereotypes” or “No evident harmful elements”).
4. **Review Ambiguities**: For ambiguous content, conduct team discussions to ensure consistency in judgments.