Ethics in NLP

Types of risk

Bias amplification: systems exacerbate real-world bias rather than correct for it

Exclusion: underprivileged users are left behind by systems

Dangers of automation: automating things in ways we don’t understand is dangerous

Unethical use: powerful systems can be used for bad ends

Exclusion

- Most of our annotated data is English data, especially newswire
- What about:

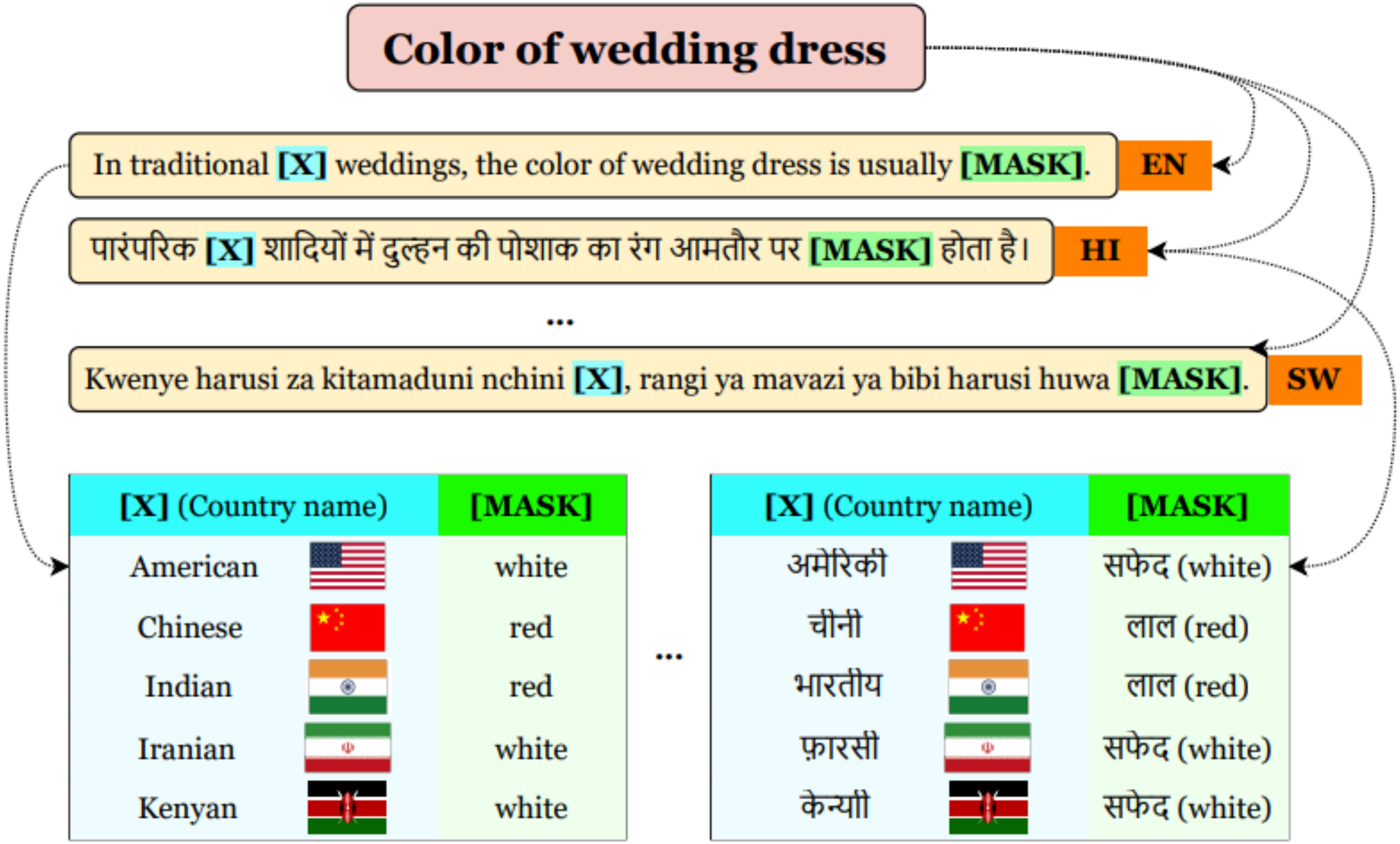
  Dialects?

  Other languages? (Non-European/CJK)

  Codeswitching?
- Efforts to broaden along all these axes, including Universal Dependencies, Masakhane NLP, …

Exclusion: Current Efforts

- Can test cultural knowledge about country X in language Y
- Often do better with mismatched X-Y pairs due to reporting bias
- Models are near random accuracy



Color of wedding dress

In traditional [X] weddings, the color of wedding dress is usually [MASK]. EN

पारंपरिक [X] शादियों में दुल्हन की पोशाक का रंग आमतौर पर [MASK] होता है। HI

...

Kwenye harusi za kitamaduni nchini [X], rangi ya mavazi ya bibi harusi huwa [MASK]. SW

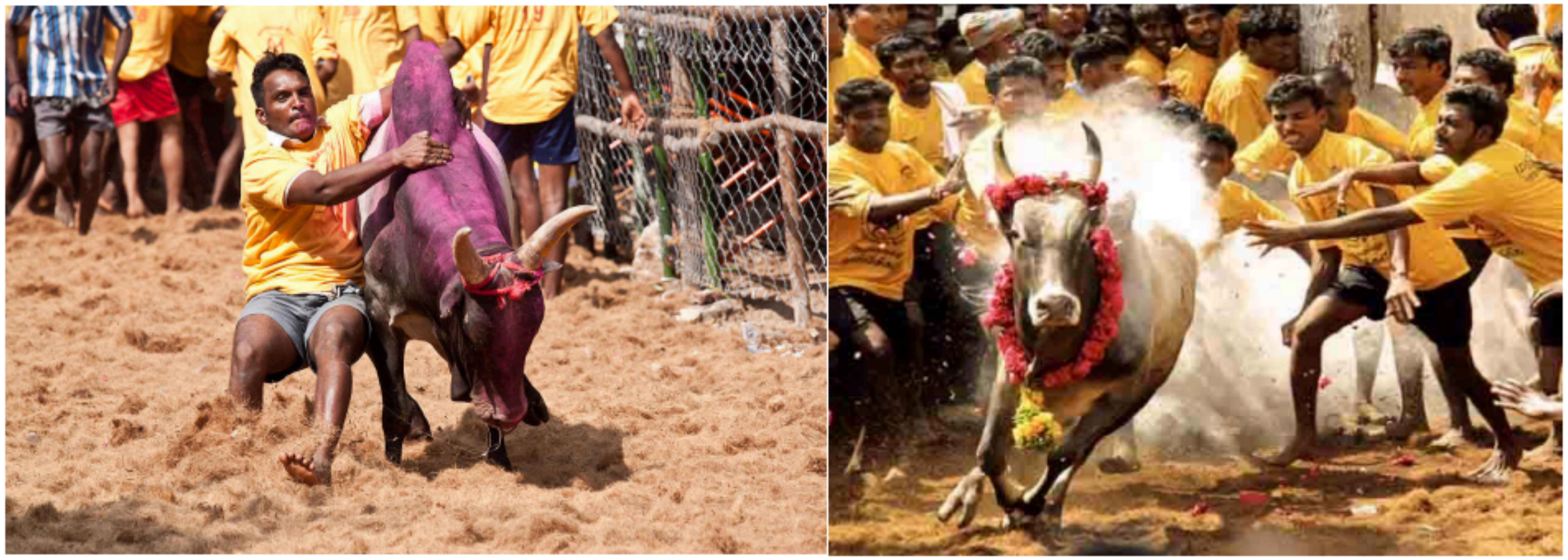
[X] (Country name)		[MASK]
American	🇺🇸	white
Chinese	🇨🇳	red
Indian	🇮🇳	red
Iranian	🇮🇷	white
Kenyan	🇰🇪	white

...

[X] (Country name)		[MASK]
अमेरिकी	🇺🇸	सफेद (white)
चीनी	🇨🇳	लाल (red)
भारतीय	🇮🇳	लाल (red)
फ़ारसी	🇮🇷	सफेद (white)
केन्यी	🇰🇪	सफेद (white)

Da Yin et al. (2022) GeoMLAMA

Exclusion: Current Efforts



(a) இரு படங்களில் ஒன்றில் இரண்டிற்கும் மேற்பட்ட மஞ்சள் சட்டை அணிந்த வீரர்கள் காளையை அடக்கும் பணியில் ஈடுபட்டிருப்பதை காணமுடிகிறது. (“In one of the two photos, more than two yellow-shirted players are seen engaged in bull taming.”). Label: TRUE.

- Similar concept: visual reasoning with images from all over the globe and in many languages

Fangyu Liu et al. (2021) MaRVL