An Image Speaks a Thousand Words, But Can Everyone Listen?
On Image Transcreation For Cultural Relevance

Simran Khanuja (and several fantastic collaborators)
Carnegie Mellon University

New Task

Language Technologies Institute

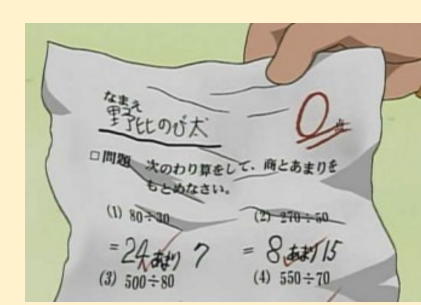## Translate **images** from one culture to another, a.k.a, **image transcreation**

1 — Real-World Applications Need MT for the Visual World
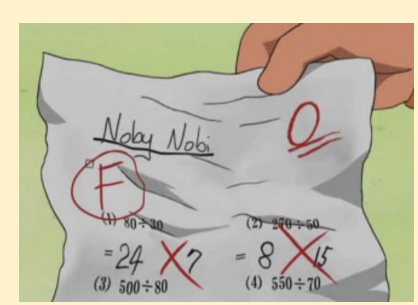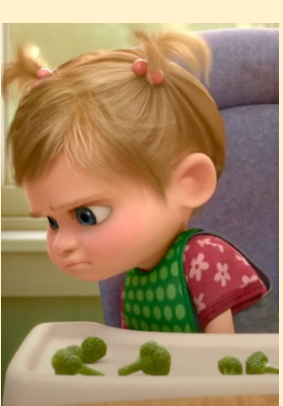
Audiovisual Media

Doremon: Japan → USA
Pixar: USA → Japan
China → Brazil
Top: Inside Out, Bottom: Zootopia

Education

Addition
USA | India

Counting
USA | India

Advertisement

Ferrero Rocher
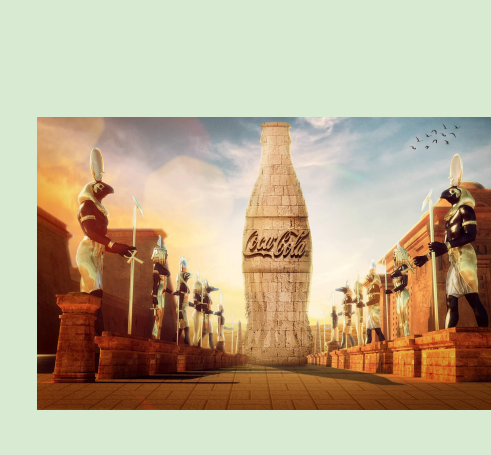China | Multiple

Coca-Cola
Multiple | Egypt

5 — Resources
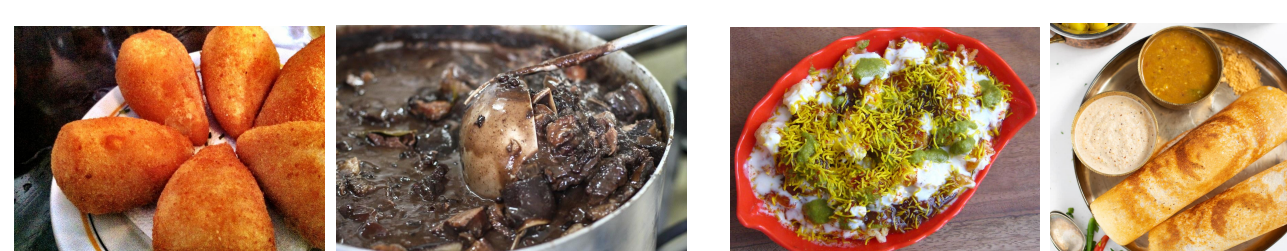
Task Paper | **CAIRe Demo**

HILITe Demo

Please reach out to skhanuja@andrew.cmu.edu for further questions!

2 — Operationalizing the Task for AI Researchers (*EMNLP '24 Best Paper* ⭐)

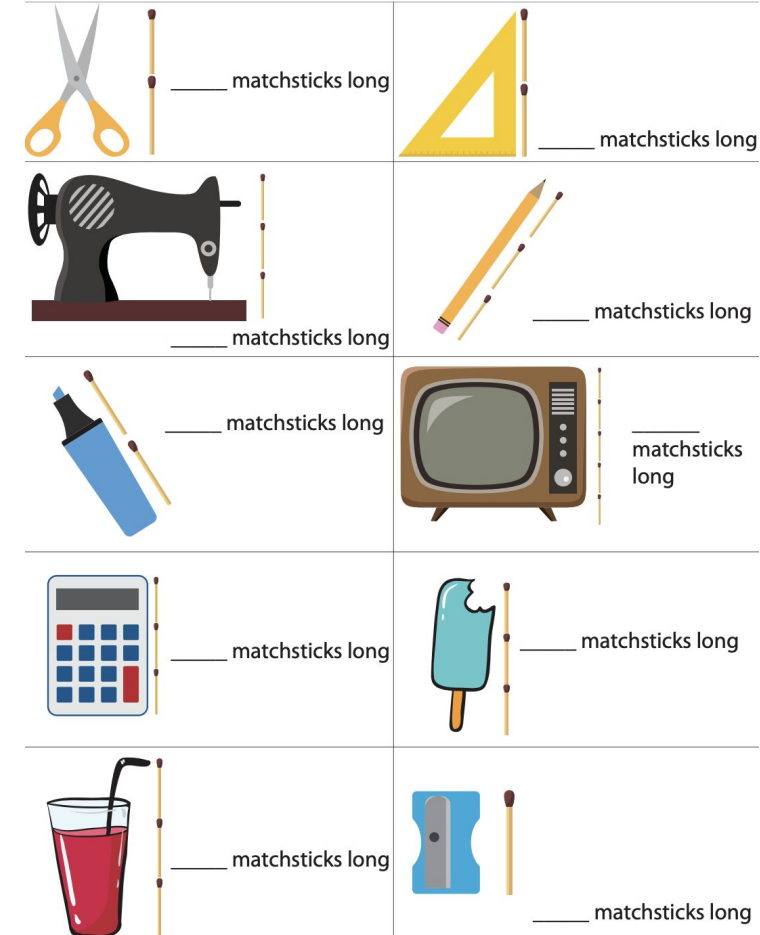**Test Set Curation**

Concept (Simple)
Brazil (food) | India (food)
Turkey (clothing) | Japan (clothing)

Application (Hard)

Evaluation Criteria
- Increase in **cultural relevance** for target
- Model output belongs to same **category**
- Increase in **cultural relevance** for target
- [Education] Same **concept** is taught
- [Stories] **Matches text** of story
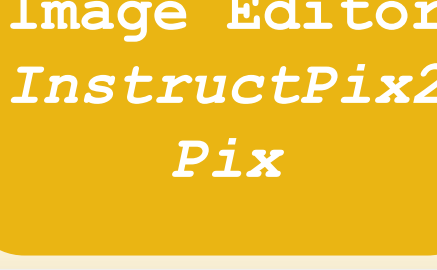
Three Pipelines

Input Image
Category: Food
Country: Nigeria
(amala-and-ewedu)
Target: United States

e2e-instruct: Instruction-based image editing
Instruction: make this image culturally relevant to the United States → Image Editor *InstructPix2Pix* →

cap-edit: Caption, text-edit, image-edit
Caption: a plate of food with meat and greens on it → LLM-edit for cultural relevance *(target: United States)*: a dish of food with beef and vegetables on it → Image Editor *PlugnPlay* →

cap-retrieve: Caption, text-edit, image-retrieval
Caption: a plate of food with meat and greens on it → LLM-edit for cultural relevance *(target: United States)*: a dish of food with beef and vegetables on it → Query LAION-US (.us in URL) →
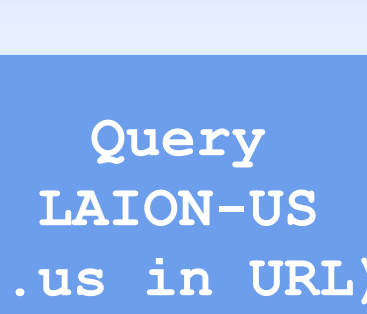
Challenges + Solutions

Lack of reliable **automatic evaluation** metrics ③

Models [mis-/under-] represent most cultures ③

No ground truth reference, subjective task ④

Applications have different desiderata, need **human-in-the-loop** ④

3 — CAIRE: Evaluating Cultural Relevance of Images

INPUT

What cultures is this image relevant to?

Countries: Afghanistan, Albania … Yemen, Zambia, Zimbabwe
Religions: Christianity, Islam, … Secular Humanism
Values: Integrity, Resilience … Prosperity, Discipline

OUTPUT

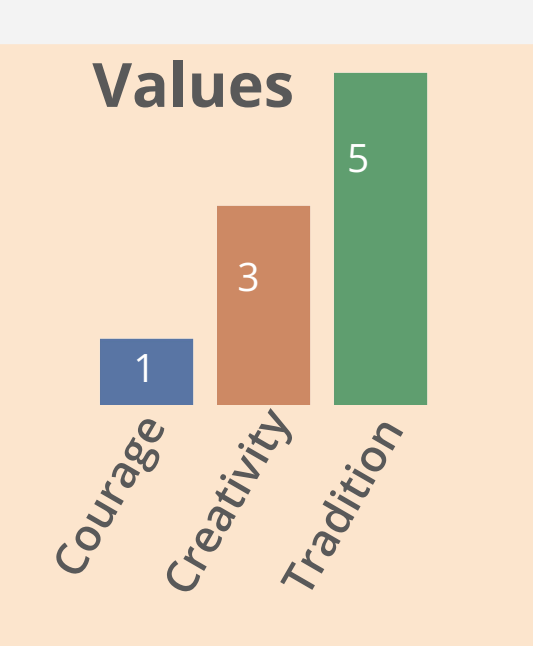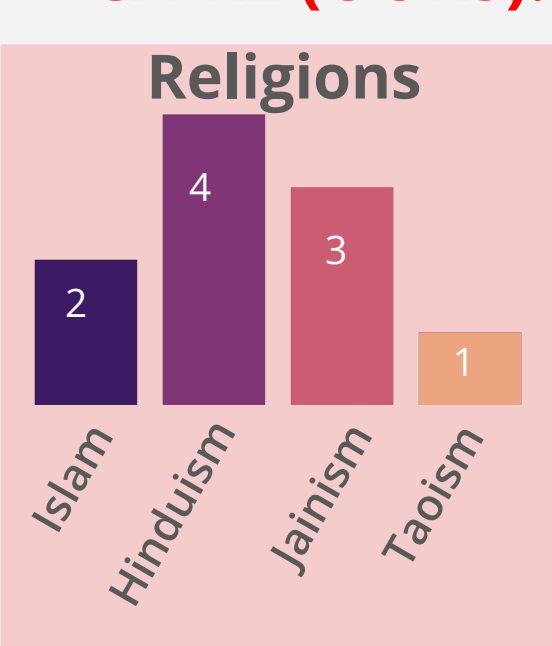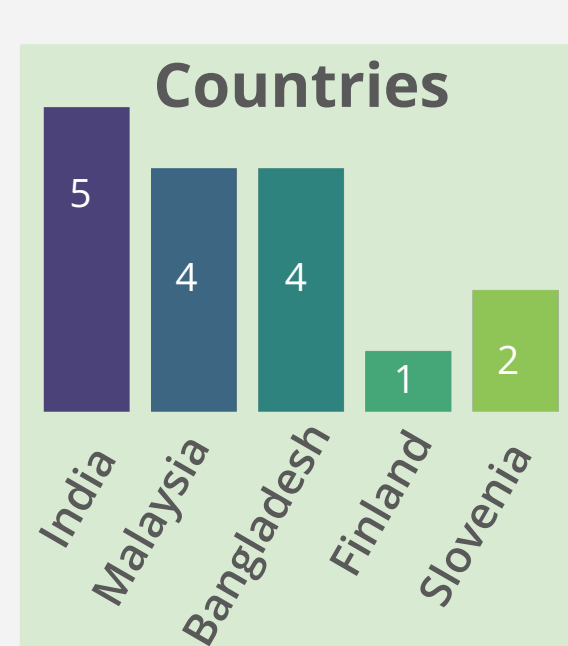**GPT4o:** South Asia, Southeast Asia, Latin America and the Caribbean

Gemini 2.0: India, Pakistan, Philippines, Mexico, Thailand, Brazil

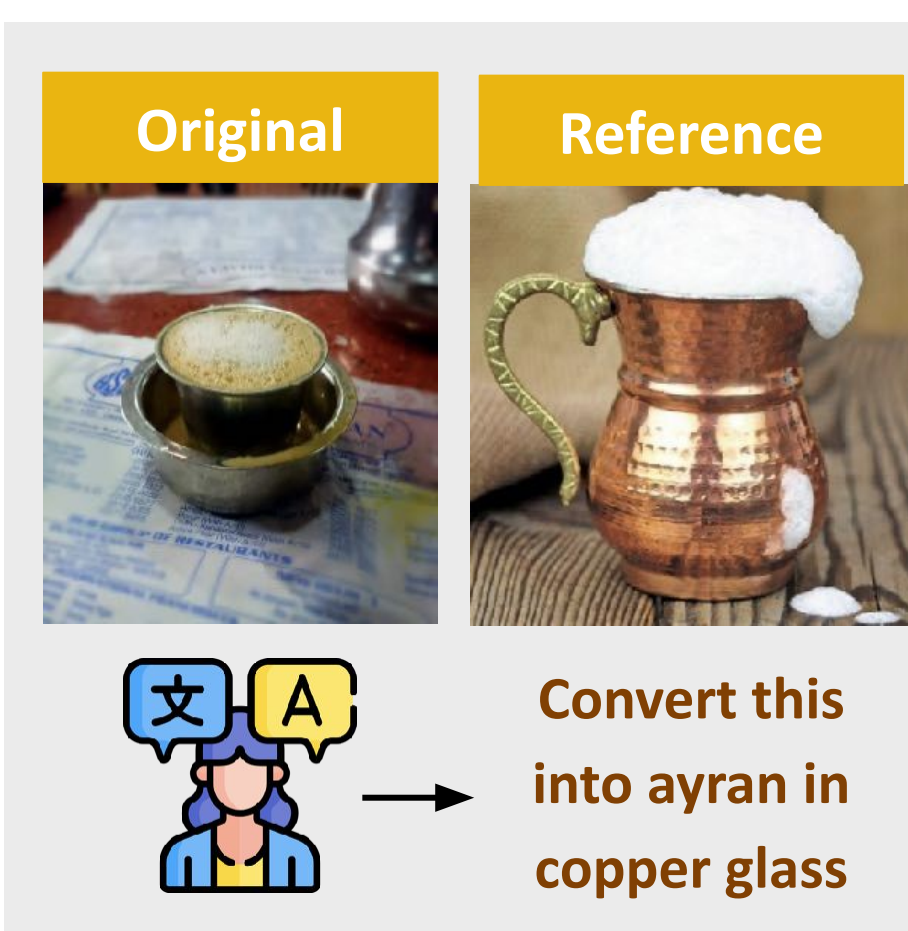**QwenVL 2.5:** India, Pakistan, Bangladesh, Thailand, Latin America, Caribbean, Africa, Australia

LLaMa 3.2: India, Southeast Asia, Caribbean, Latin America, Africa

CAIRE (OURS):
Countries: India, Malaysia, Bangladesh, Finland, Slovenia
Religions: Islam, Hinduism, Jainism, Taoism
Values: Courage, Creativity, Tradition

4 — HILITe: Human-AI Collaborative Platform for Editing

Original | Reference

Intent Detection & Run Models
- Object Detection
- Reference Image

Preferences
- Generation Quality
- Text Influence
- Image Influence

Models: AnyDoor, InstructPix2Pix, Paint-by-example, StyleCLIP, ControlNet, ScreenText Eraser, ESRGAN, DiscoFaceGAN

Convert this into ayran in copper glass
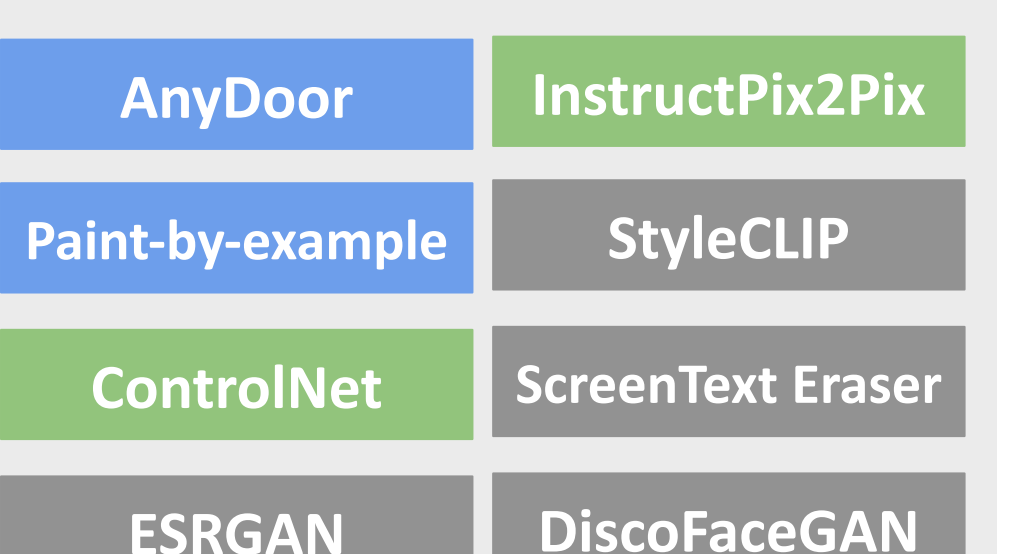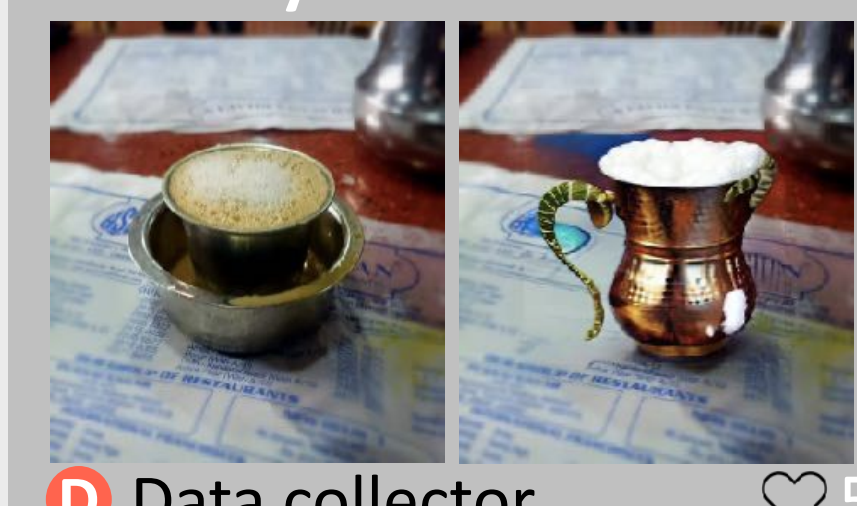
Iterate

Publish & Gallery
From: India | To: Turkey | Visibility: Public
coffee-ayran
Ⓓ Data collector ♡ 5

Feedback Ratings Collected for:
- Cultural Accuracy
- Image Quality
- Instruction Following
- Input Img Similarity