Columbia University

The University of North Carolina at Chapel Hill

See It From My Perspective:

How Language Affects Cultural Bias in Image Understanding

Amith Ananthram, Elias Stengel-Eskin, Mohit Bansal, Kathleen McKeown

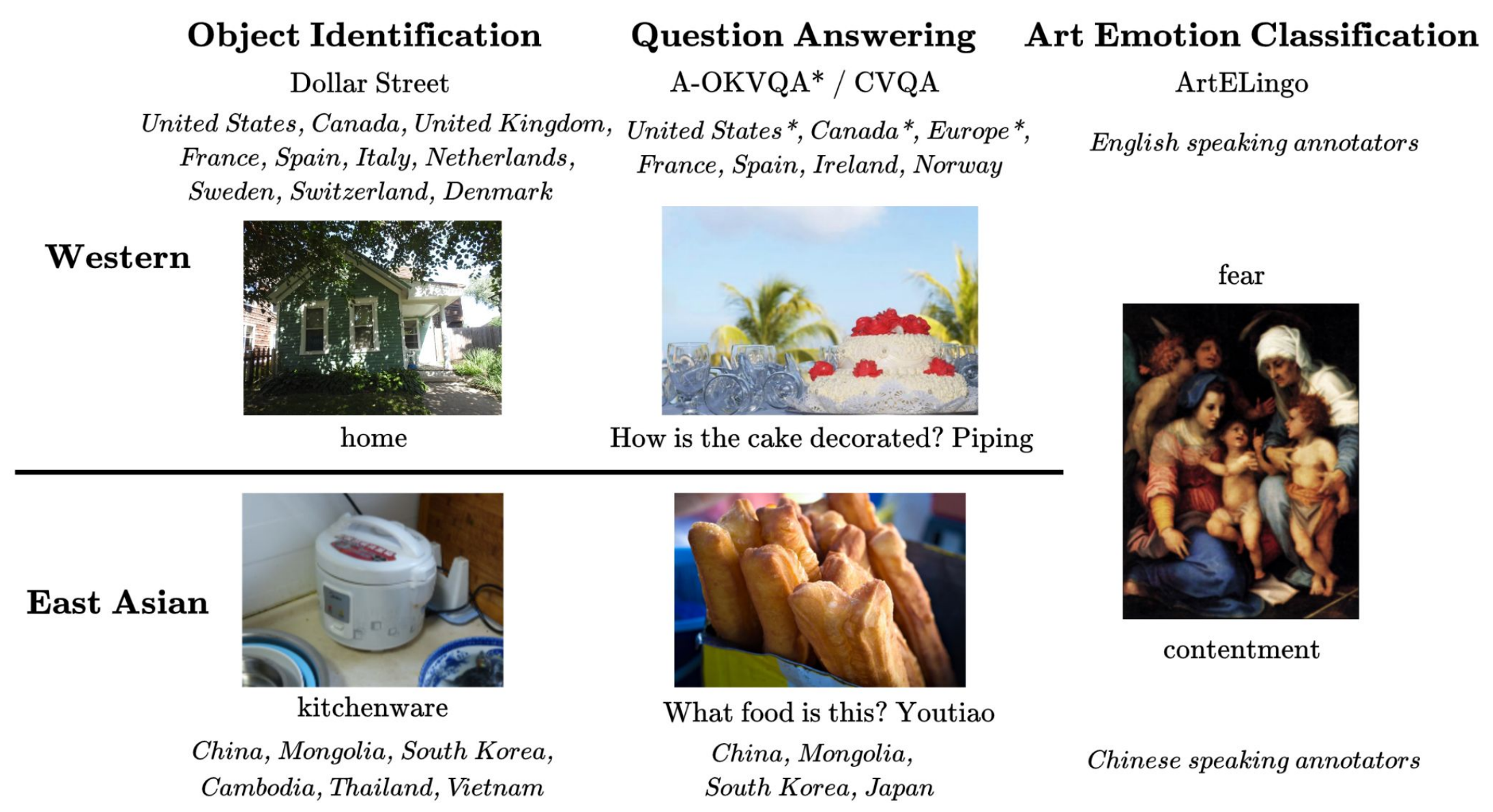
1. Background

- our knowledge and our beliefs are informed by culture (Goldstein, 1957)
- culture affects *how* we see things [color grouping (Chiao & Harada, 2008); attentional focus (Nisbett, 2001)]
- LLMs exhibit a Western worldview in knowledge & beliefs (Xu, 2024)
- VLMs inherit knowledge (Tsimpoukelli, 2021) and multilinguality from their LLMs

2. Our Questions

1) **Whose perspective** do VLMs model?

2) How does using a **culturally closer language** affect this perspective?



Object Identification
Dollar Street
United States, Canada, United Kingdom, France, Spain, Italy, Netherlands, Sweden, Switzerland, Denmark

Question Answering
A-OKVQA* / CVQA
United States, Canada*, Europe*, France, Spain, Ireland, Norway*

Art Emotion Classification
ArtELingo
English speaking annotators

Western

home

How is the cake decorated? Piping

fear

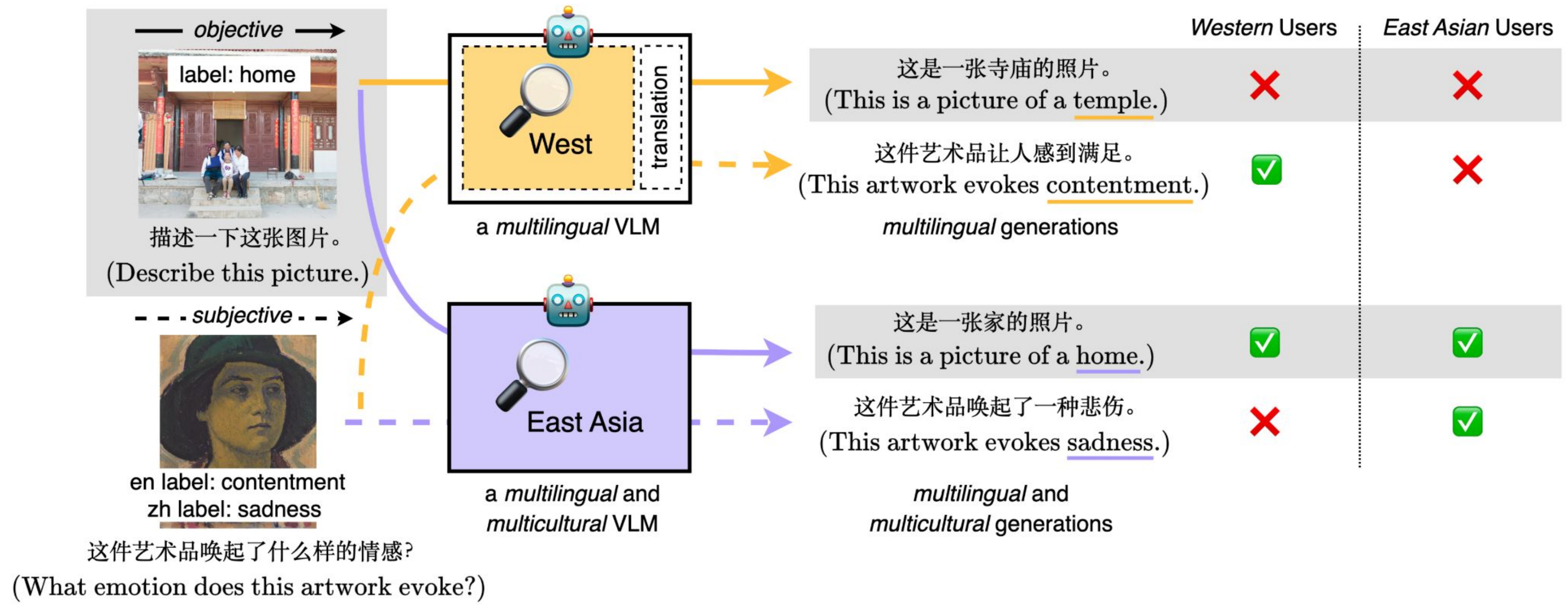
East Asian

kitchenware
China, Mongolia, South Korea, Cambodia, Thailand, Vietnam

What food is this? Youtiao
China, Mongolia, South Korea, Japan

contentment
Chinese speaking annotators

3. Approach



objective

label: home

描述一下这张图片。
(Describe this picture.)

subjective

en label: contentment
zh label: sadness

这件艺术品唤起了什么样的情感?
(What emotion does this artwork evoke?)

West

translation

a *multilingual* VLM

East Asia

a *multilingual* and *multicultural* VLM

Western Users | *East Asian* Users

这是一张寺庙的照片。
(This is a picture of a temple.) ✗ ✗

这件艺术品让人感到满足。
(This artwork evokes contentment.) ✓ ✗

multilingual generations

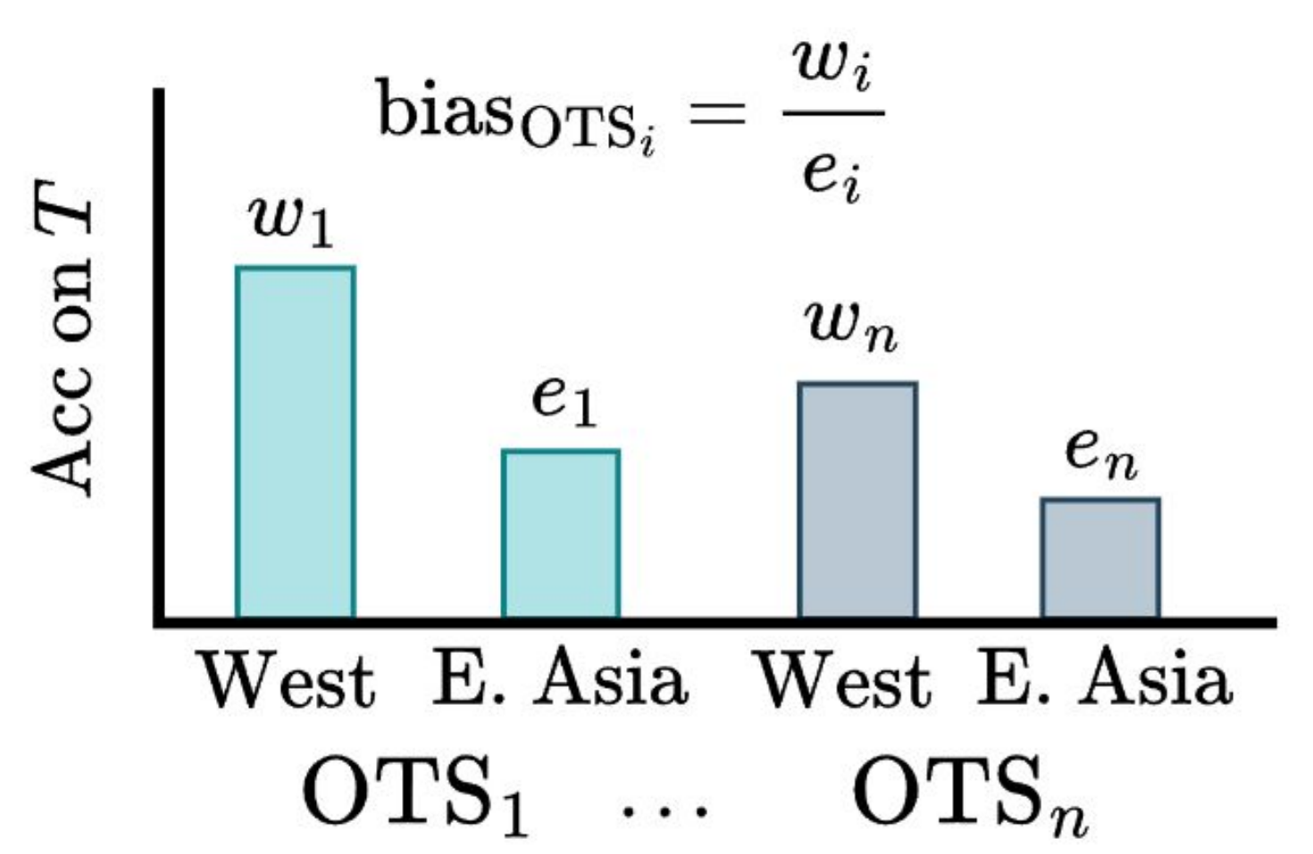
这是一张家的照片。
(This is a picture of a home.) ✓ ✓

这件艺术品唤起了一种悲伤。
(This artwork evokes sadness.) ✗ ✓

multilingual and *multicultural* generations

Step 1: Bias Characterization

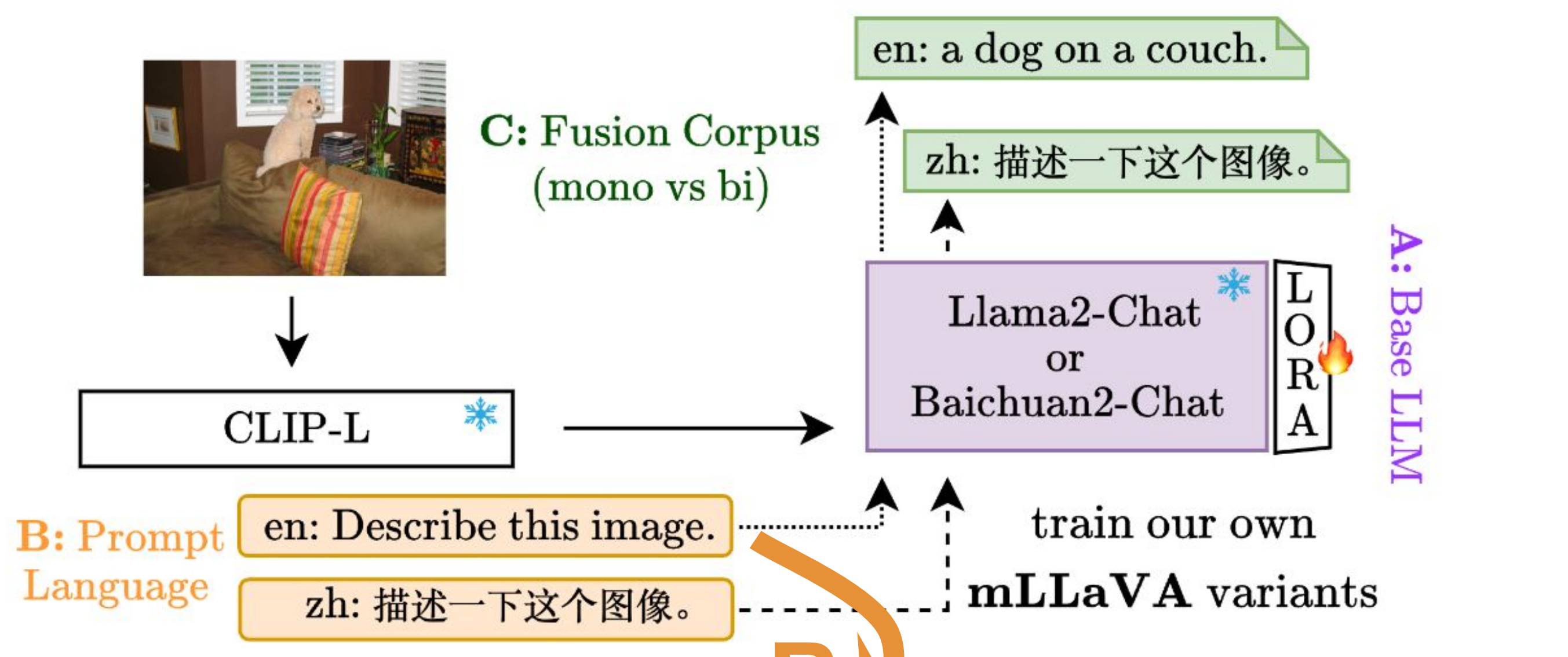
Do off-the-shelf (OTS_i) VLMs have a Western bias?



$$\text{bias}_{\text{OTS}_i} = \frac{w_i}{e_i}$$

Step 2: Bias Sourcing

How does language affect Western bias in VLMs?



C: Fusion Corpus (mono vs bi)

en: a dog on a couch.

zh: 描述一下这个图像。

CLIP-L

Llama2-Chat or Baichuan2-Chat

LORA

A: Base LLM

B: Prompt Language

en: Describe this image.

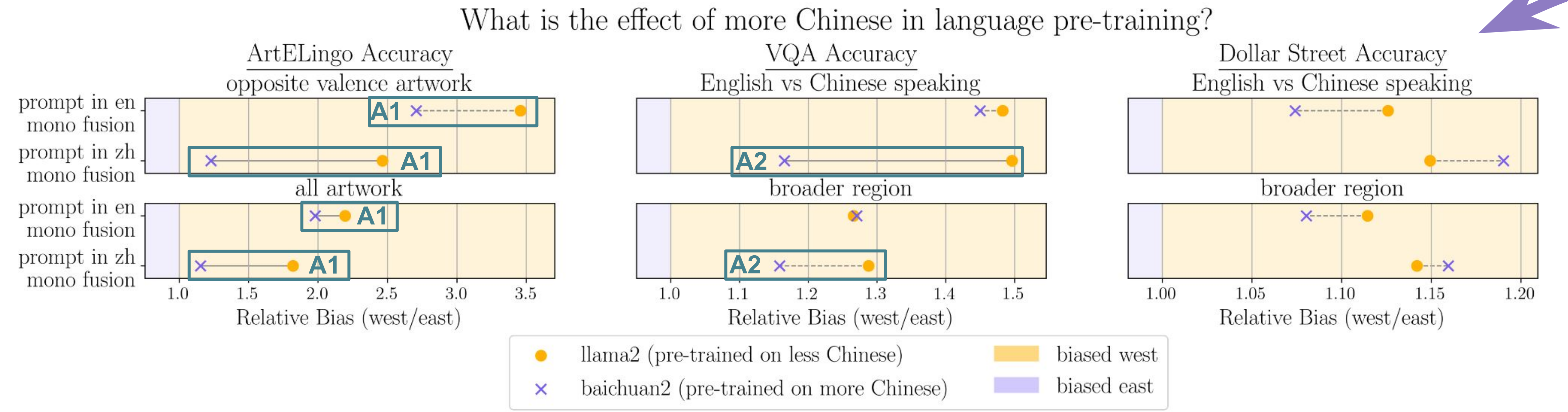
zh: 描述一下这个图像。

train our own **mLLaVA** variants

B

4. Results

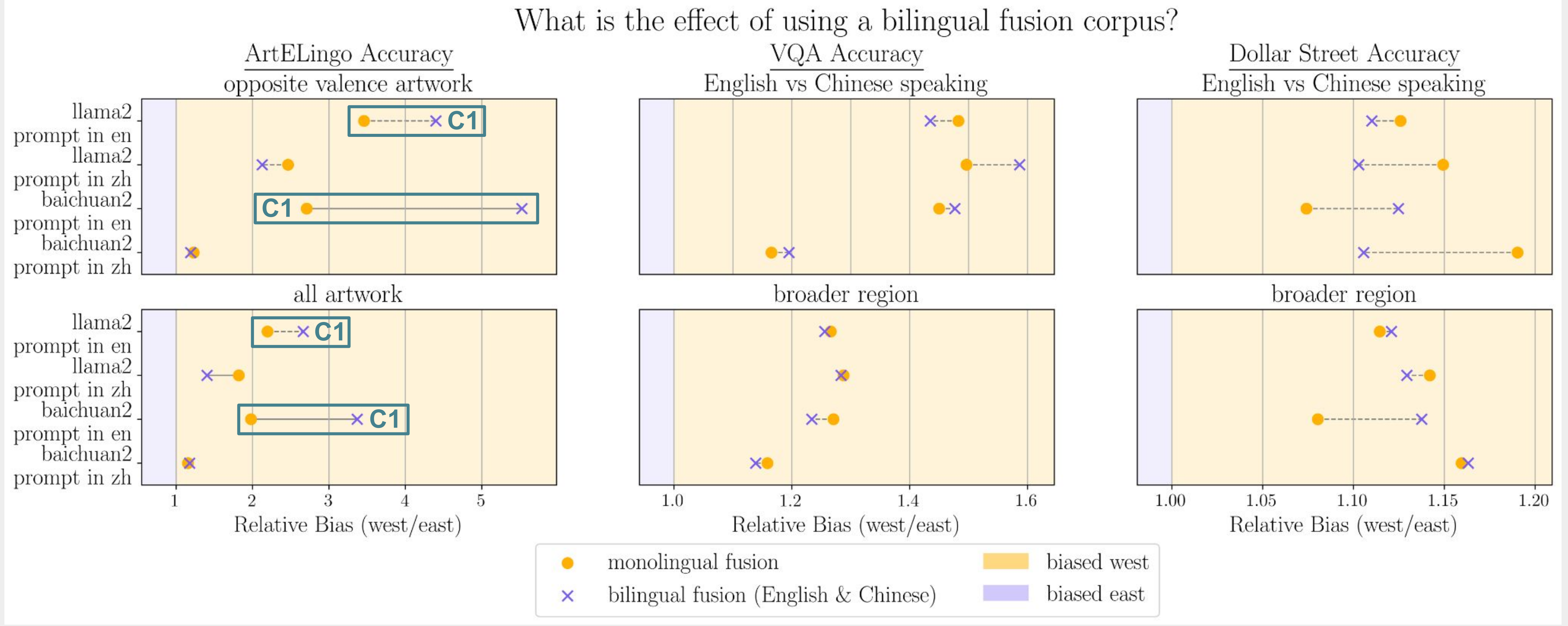
A



What is the effect of more Chinese in language pre-training?

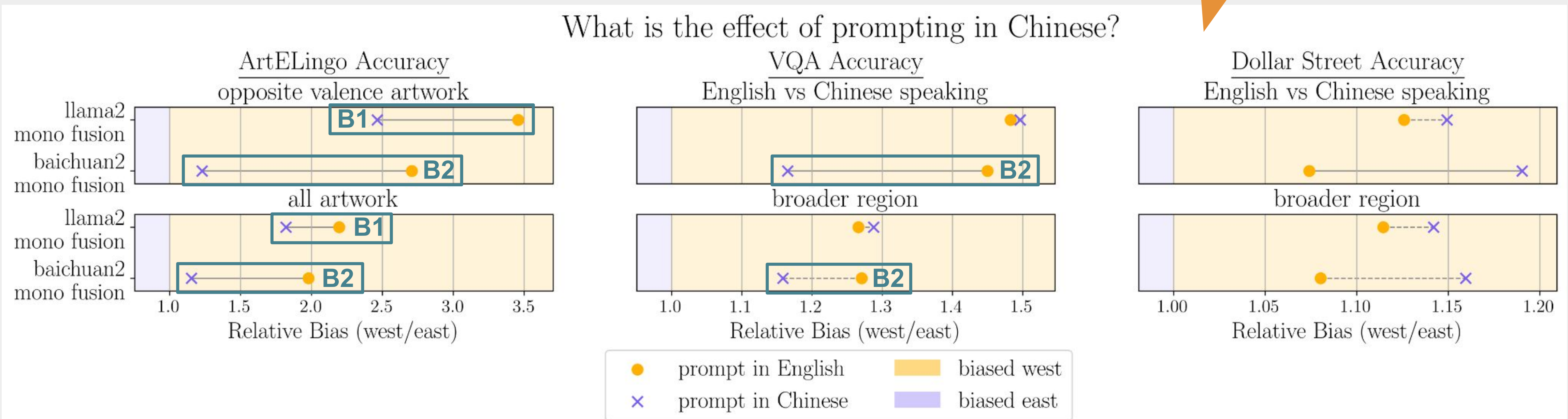
A1. On *subjective* tasks, it reduces bias when prompting in both *English* and *Chinese.*
A2. On *objective* tasks, it reduces bias when prompting in *Chinese*.

C



What is the effect of using a bilingual fusion corpus?

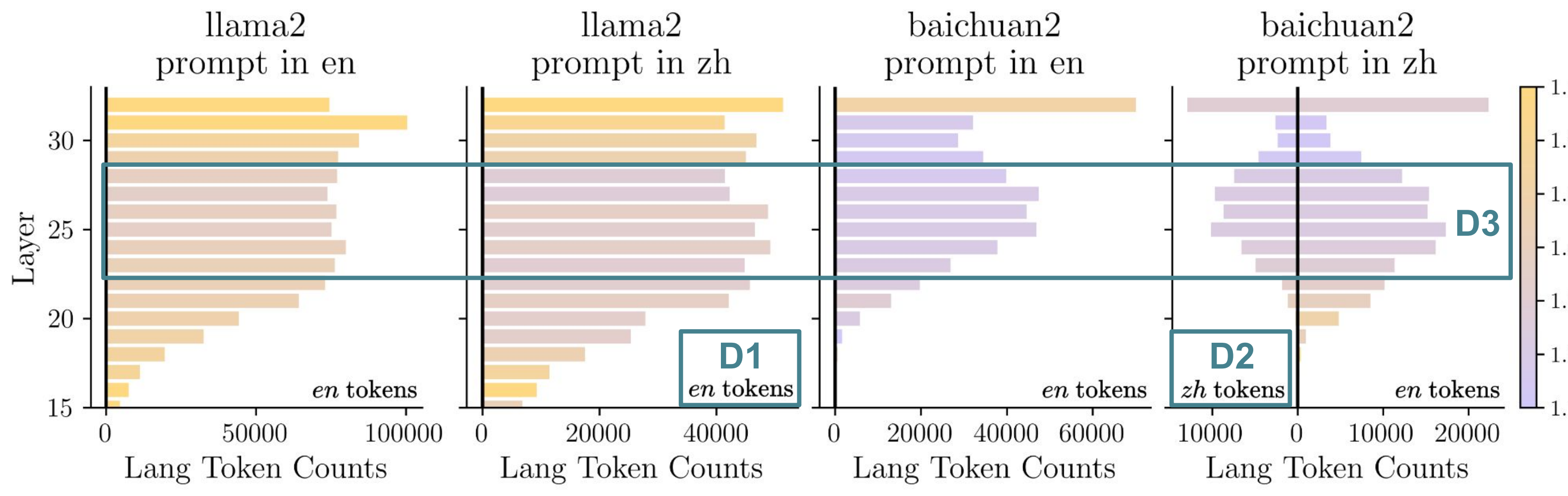
C1. On *subjective* tasks, it ties a language to its speakers' perspective (esp. in English).



What is the effect of prompting in Chinese?

B1. On *subjective* tasks, it reduces bias *even if* little Chinese was seen during pre-training.
B2. However, it is more effective when Chinese was *common during pre-training.*

Mechanistic Exploration of Hidden States



D1. In Llama2, decode to English.

D2. In Baichuan2, decode to English and Chinese.

D3. In Baichuan2, *more similar* to rationales from Chinese language annotators than in Llama2.

Takeaway:

A prompt language *can* reduce cultural bias in VLMs but the text-only pre-training language mix matters more; MT / bilingual fusion are insufficient proxies.