CS329X: Human Centered NLP

Culture and Values

Diyi Yang

Stanford CS

Lecture Overview

- Cross-culture NLP
  Slides credit to Daniel Hershcovich, EACL workshop on CrossCultural Considerations in NLP
- Value Alignment in LLMs
- Values in ML Research

Slides credits to:
- Camille Harris on Values in ML Research
- Daniel Hershcovich on Cross-culture NLP

What is culture?

shared beliefs, values, practices, behaviors, and artifacts

a system of **symbols**

characterizes a **social** group

learned, not innate

integrated with daily **experience**

taken-for-granted notions, rules, moralities, and behaviors

dynamic **process** through which social meanings are constructed

Cultural anthropology: Asking questions about humanity (Welsch and Vivanco, 2015)

Take a closer look at culture

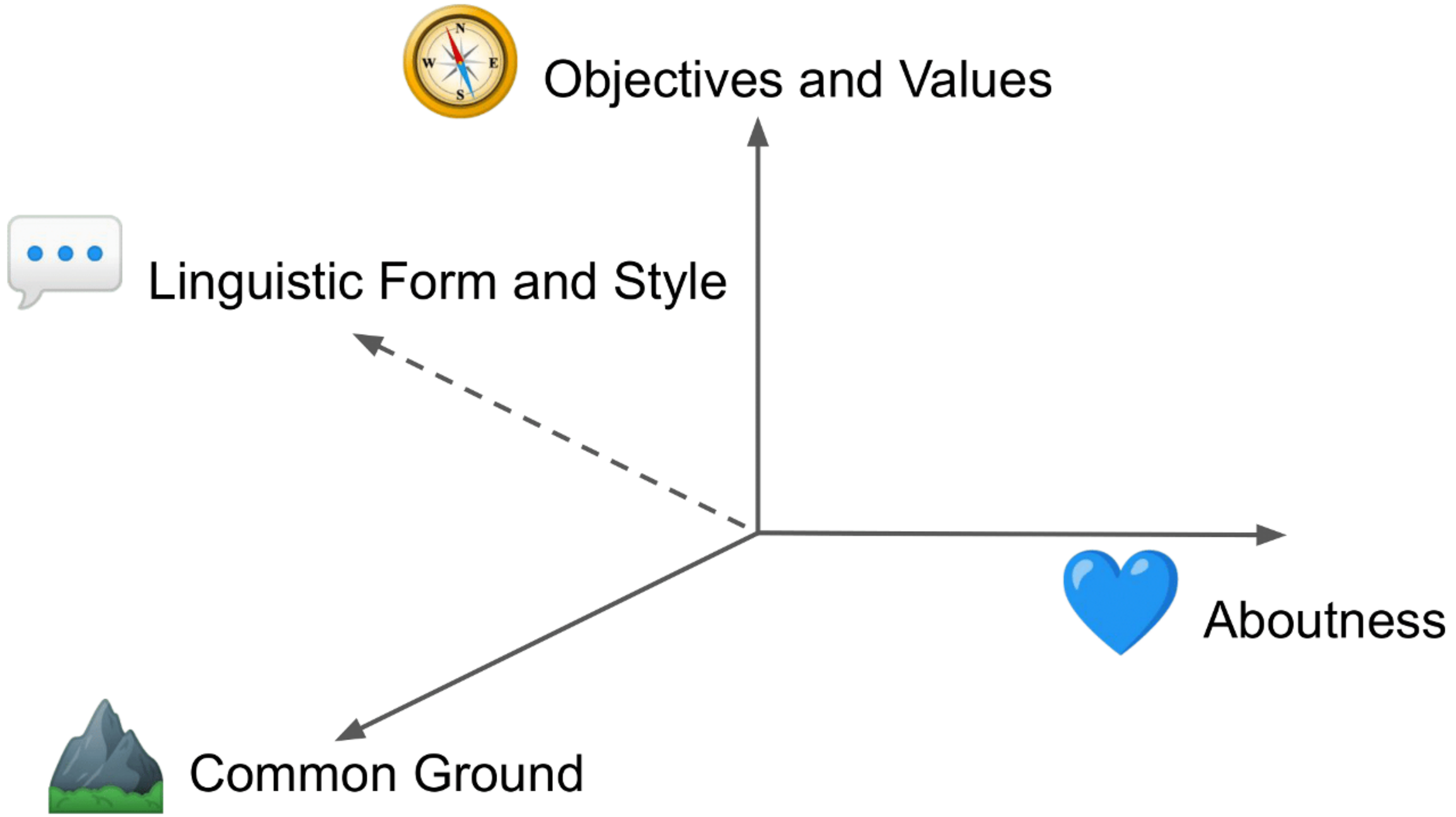
Linguistic variation

- Idiolect
- Sociolect
- Dialect
- Standardized language
- Language groups and families

Cultural variation

- Individual, personality
- Social group
- Geographical region
- Country, national culture
- Global cultural community

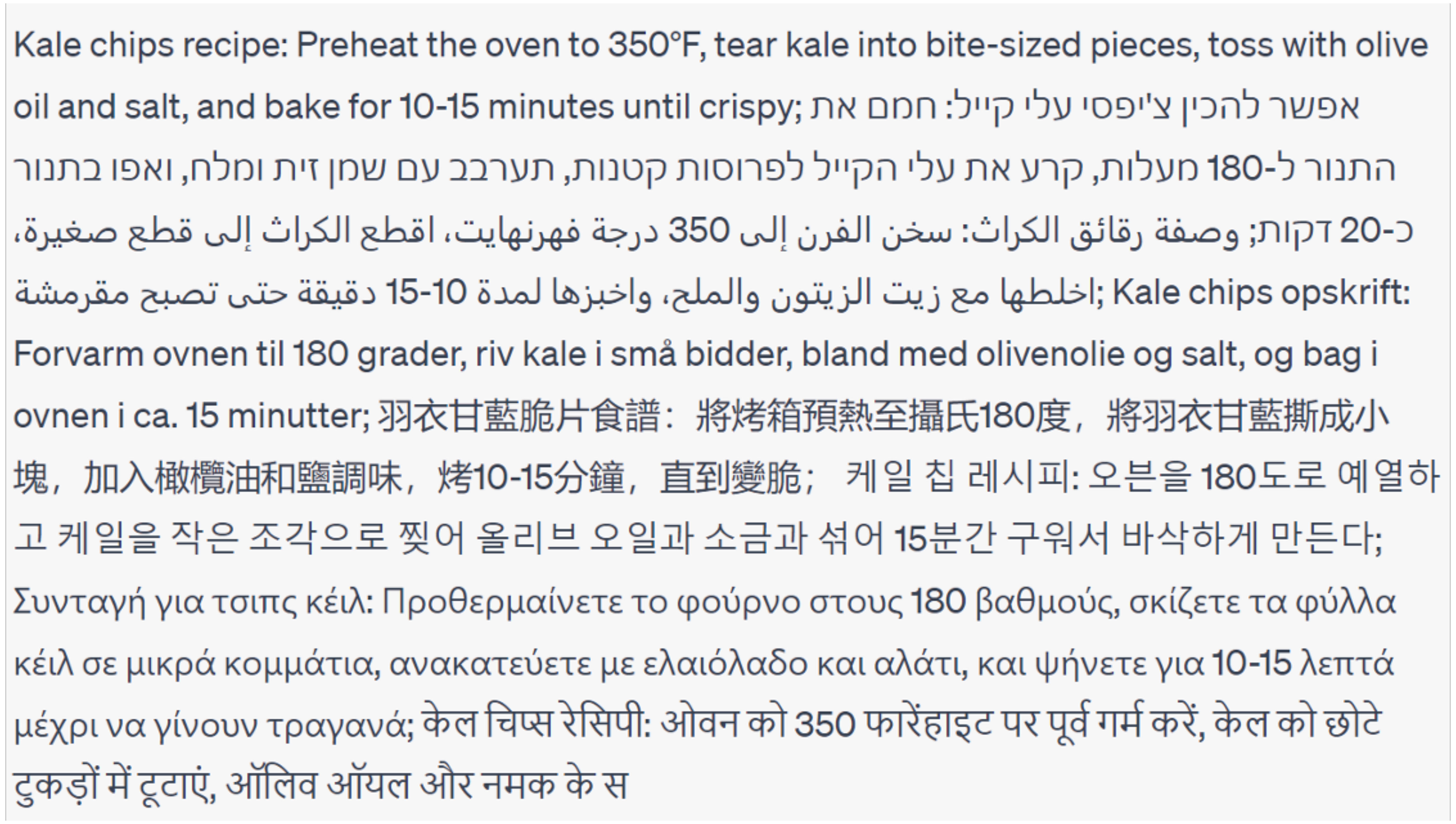
Cultural dimensions in NLP



Challenges and Strategies in Cross-Cultural NLP (Hershcovich et al., ACL 2022)

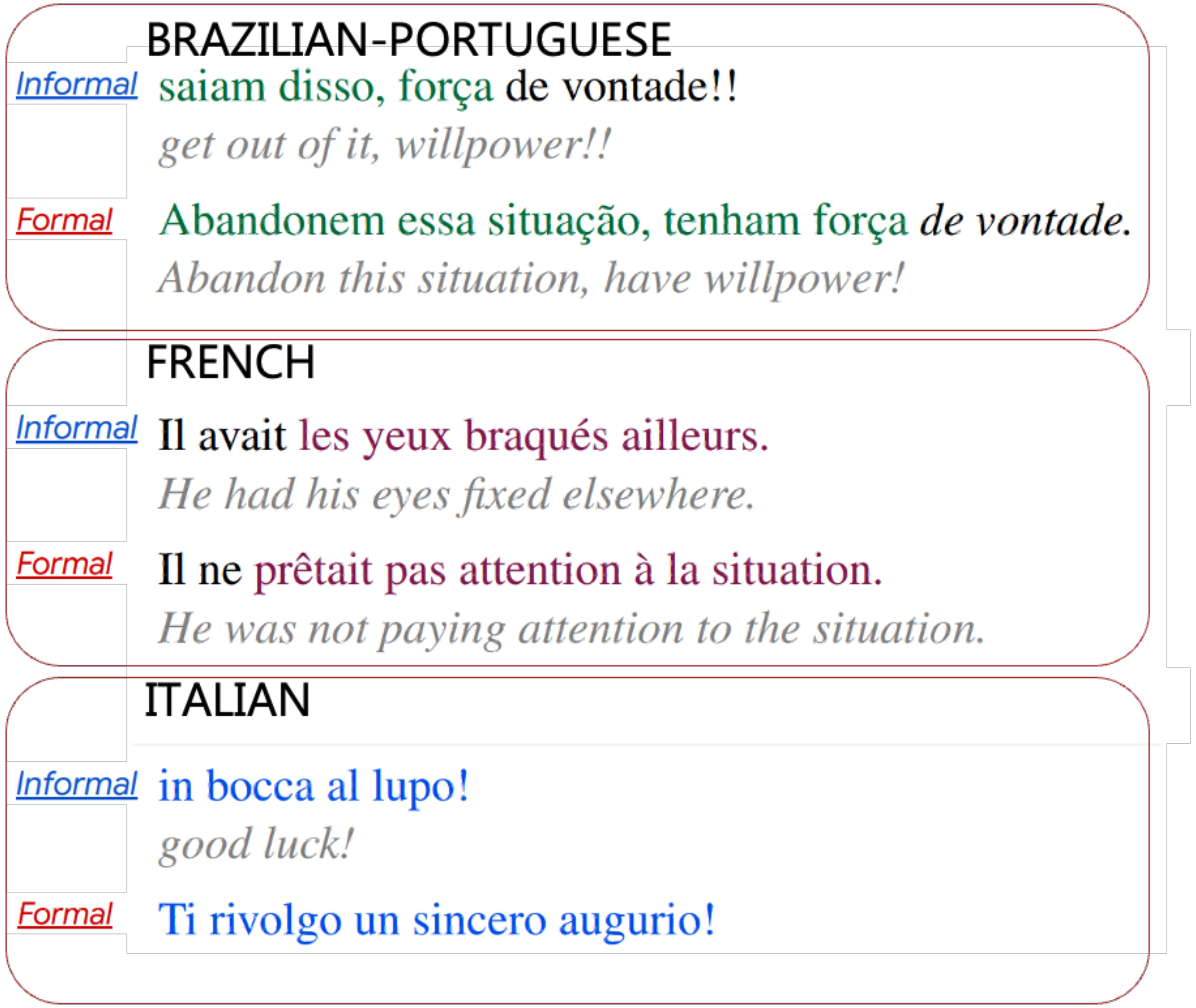
Linguistic form

Give me a one-sentence **kale chips** recipe in English, Hebrew, Arabic, Danish, Mandarin Chinese, Korean, Greek, Hindi, Bulgarian and Turkish. Do not say which languages they are.



Kale chips recipe: Preheat the oven to 350°F, tear kale into bite-sized pieces, toss with olive oil and salt, and bake for 10-15 minutes until crispy; אפשר להכין צ'יפסי עלי קייל: חמם את התנור ל-180 מעלות, קרע את עלי הקייל לפרוסות קטנות, תערבב עם שמן זית ומלח, ואפו בתנור כ-20 דקות; وصفة رقائق الكراث: سخن الفرن إلى 350 درجة فهرنهايت، اقطع الكراث إلى قطع صغيرة، اخلطها مع زيت الزيتون والملح، واخبزها لمدة 15-10 دقيقة حتى تصبح مقرمشة; Kale chips opskrift: Forvarm ovnen til 180 grader, riv kale i små bidder, bland med olivenolie og salt, og bag i ovnen i ca. 15 minutter; 羽衣甘藍脆片食譜：將烤箱預熱至攝氏180度，將羽衣甘藍撕成小塊，加入橄欖油和鹽調味，烤10-15分鐘，直到變脆； 케일 칩 레시피: 오븐을 180도로 예열하고 케일을 작은 조각으로 찢어 올리브 오일과 소금과 섞어 15분간 구워서 바삭하게 만든다; Συνταγή για τσιπς κέιλ: Προθερμαίνετε το φούρνο στους 180 βαθμούς, σκίζετε τα φύλλα κέιλ σε μικρά κομμάτια, ανακατεύετε με ελαιόλαδο και αλάτι, και ψήνετε για 10-15 λεπτά μέχρι να γίνουν τραγανά; केल चिप्स रेसिपी: ओवन को 350 फारेंहाइट पर पूर्व गर्म करें, केल को छोटे टुकड़ों में टूटाएं, ऑलिव ऑयल और नमक के स

Style



BRAZILIAN-PORTUGUESE

Informal saiam disso, força de vontade!!

get out of it, willpower!!

Formal Abandonem essa situação, tenham força *de vontade.*

Abandon this situation, have willpower!

FRENCH

Informal Il avait les yeux braqués ailleurs.

He had his eyes fixed elsewhere.

Formal Il ne prêtait pas attention à la situation.

He was not paying attention to the situation.

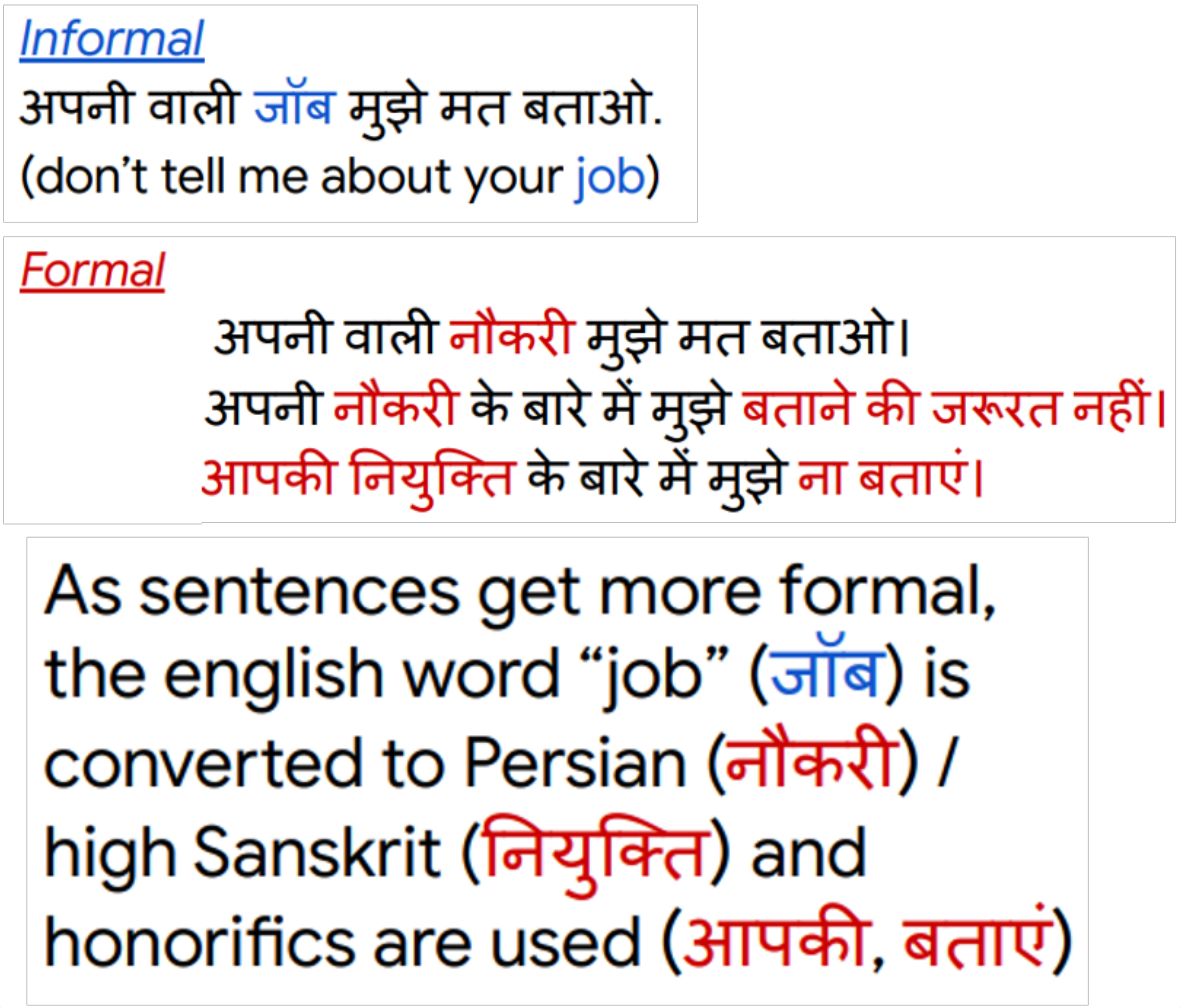
ITALIAN

Informal in bocca al lupo!

good luck!

Formal Ti rivolgo un sincero augurio!

Olá, Bonjour, Salve! XFORMAL: A Benchmark for Multilingual Formality Style Transfer (Briakou et al., NAACL 2021)



Informal

अपनी वाली जॉब मुझे मत बताओ.

(don’t tell me about your job)

Formal

अपनी वाली नौकरी मुझे मत बताओ।

अपनी नौकरी के बारे में मुझे बताने की जरूरत नहीं।

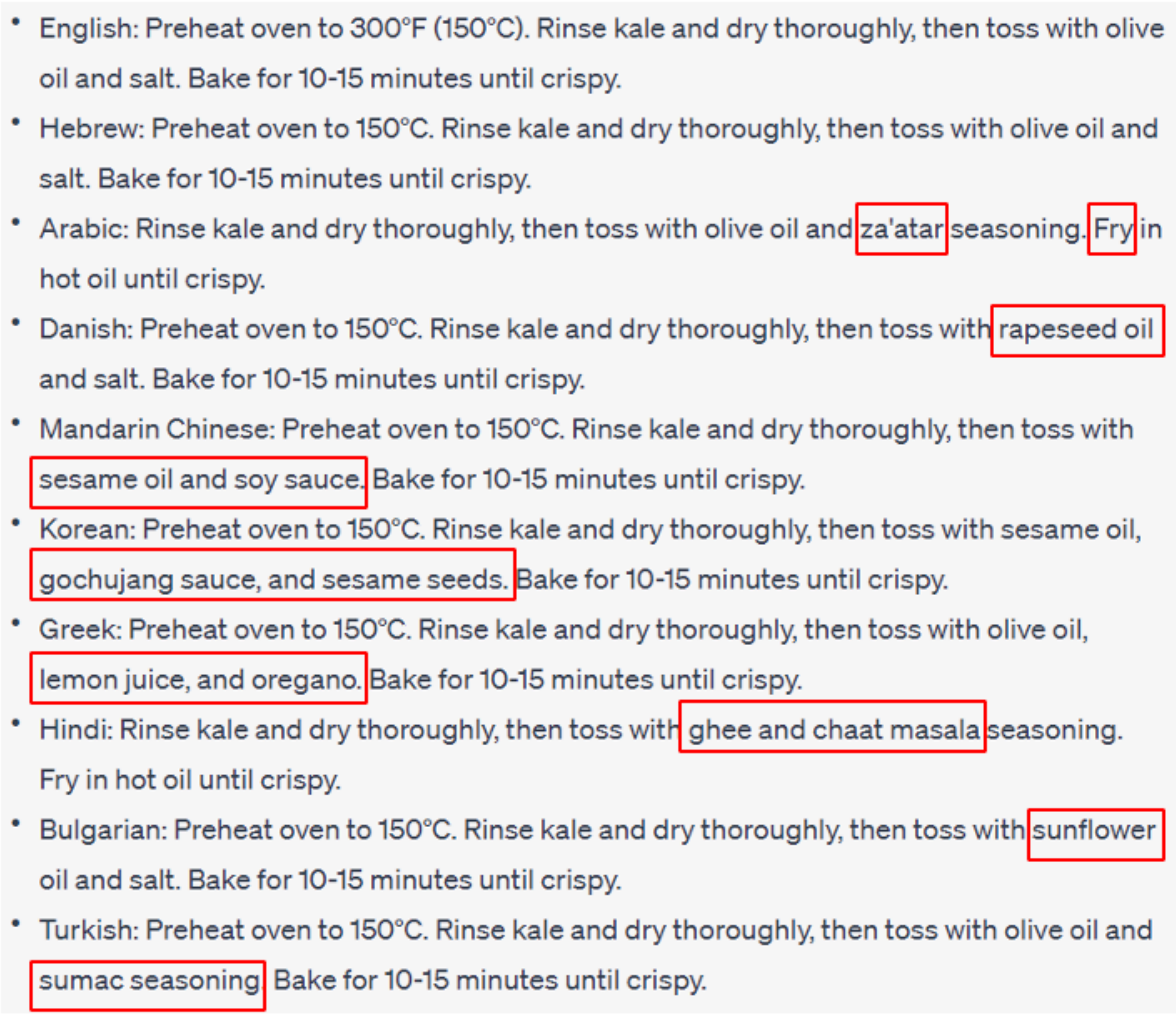
आपकी नियुक्ति के बारे में मुझे ना बताएं।

As sentences get more formal, the english word “job” (जॉब) is converted to Persian (नौकरी) / high Sanskrit (नियुक्ति) and honorifics are used (आपकी, बताएं)

Few-shot Controllable Style Transfer for Low-Resource Multilingual Settings (Krishna et al., ACL 2022)

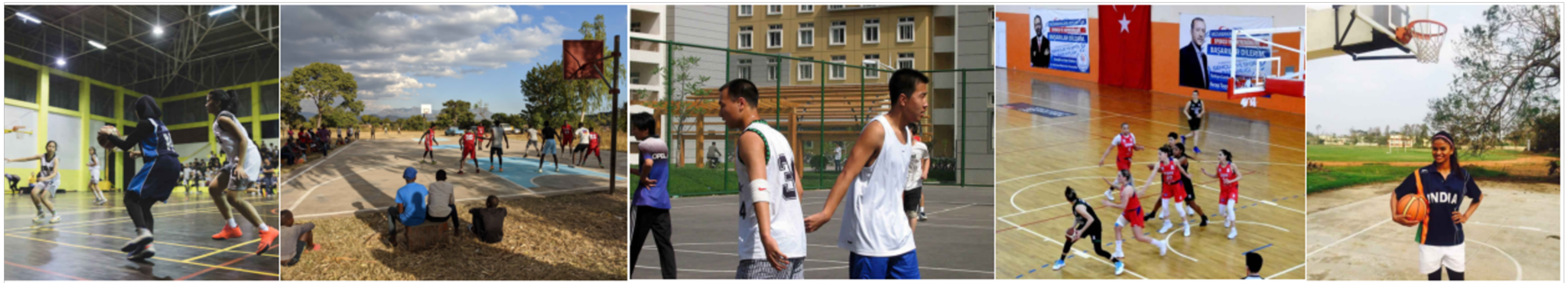
Common ground

Adapt the recipe to the **common ingredients, cooking methods and tools** in each culture's cuisine

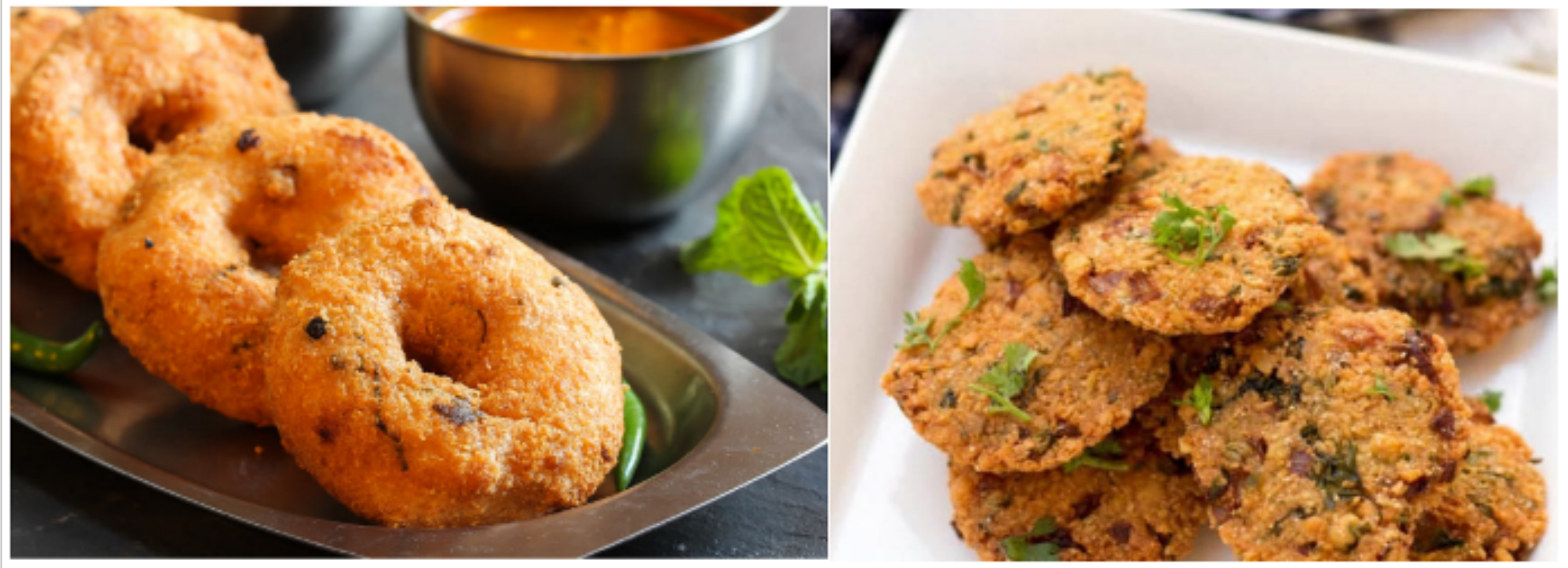


- English: Preheat oven to 300°F (150°C). Rinse kale and dry thoroughly, then toss with olive oil and salt. Bake for 10-15 minutes until crispy.
- Hebrew: Preheat oven to 150°C. Rinse kale and dry thoroughly, then toss with olive oil and salt. Bake for 10-15 minutes until crispy.
- Arabic: Rinse kale and dry thoroughly, then toss with olive oil and za'atar seasoning. Fry in hot oil until crispy.
- Danish: Preheat oven to 150°C. Rinse kale and dry thoroughly, then toss with rapeseed oil and salt. Bake for 10-15 minutes until crispy.
- Mandarin Chinese: Preheat oven to 150°C. Rinse kale and dry thoroughly, then toss with sesame oil and soy sauce. Bake for 10-15 minutes until crispy.
- Korean: Preheat oven to 150°C. Rinse kale and dry thoroughly, then toss with sesame oil, gochujang sauce, and sesame seeds. Bake for 10-15 minutes until crispy.
- Greek: Preheat oven to 150°C. Rinse kale and dry thoroughly, then toss with olive oil, lemon juice, and oregano. Bake for 10-15 minutes until crispy.
- Hindi: Rinse kale and dry thoroughly, then toss with ghee and chaat masala seasoning. Fry in hot oil until crispy.
- Bulgarian: Preheat oven to 150°C. Rinse kale and dry thoroughly, then toss with sunflower oil and salt. Bake for 10-15 minutes until crispy.
- Turkish: Preheat oven to 150°C. Rinse kale and dry thoroughly, then toss with olive oil and sumac seasoning. Bake for 10-15 minutes until crispy.

Conceptualization



Bola basket (Indonesian) *Mpira wa kikapu* (Swahili) 篮球 (Chinese) *Basketbol* (Turkish) கூடைப்பந்தாட்டம் (Tamil)

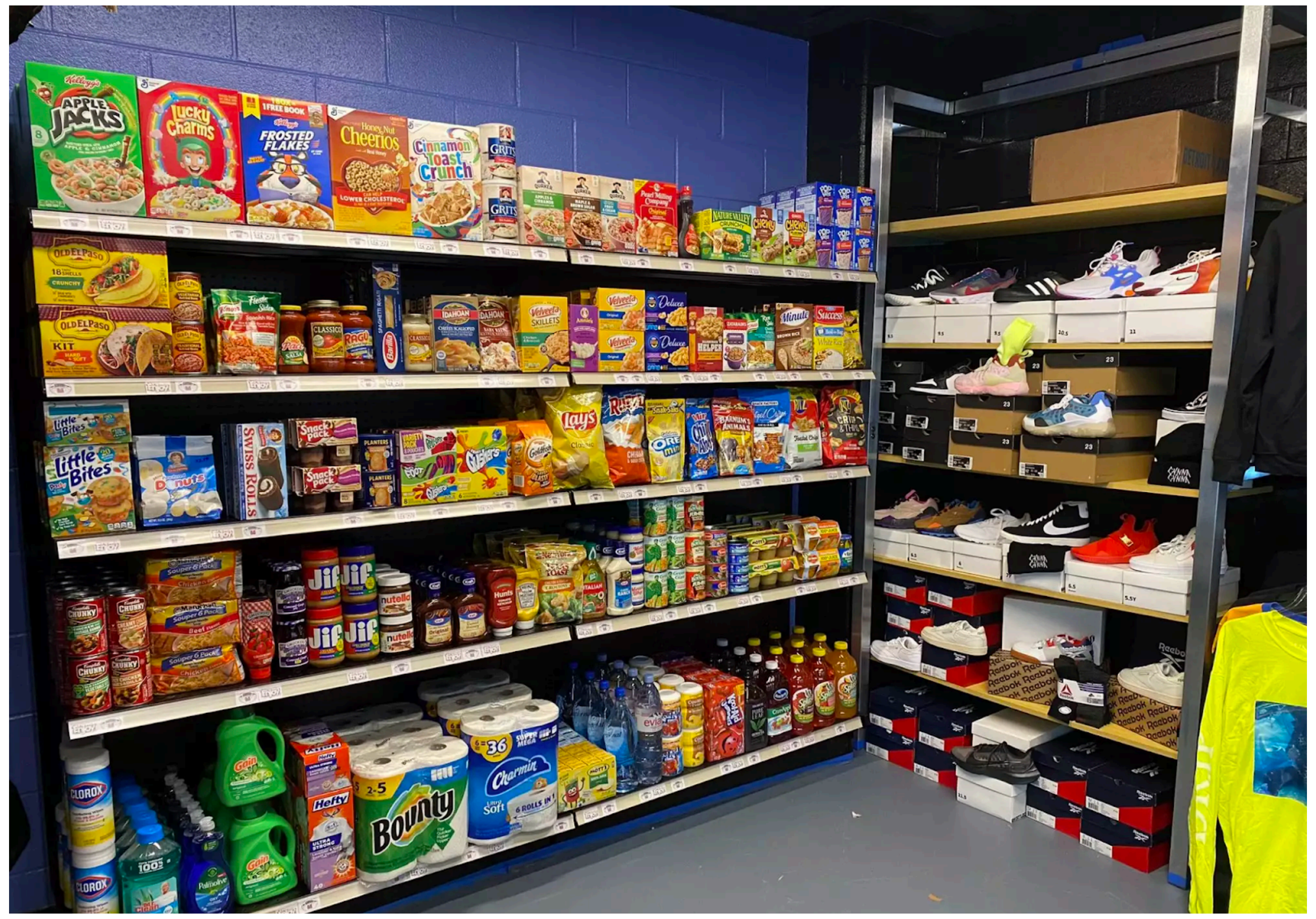


(a) இரண்டு படங்களிலும் நிறைய மசால் வடைகள் உள்ளன. (“Both images contain a lot of masala vadas.”, concept: வடை (VADA, a popular Indian food), label: FALSE)

[Visually Grounded Reasoning across Languages and Cultures](#) (Liu et al., EMNLP 2021)

Snack

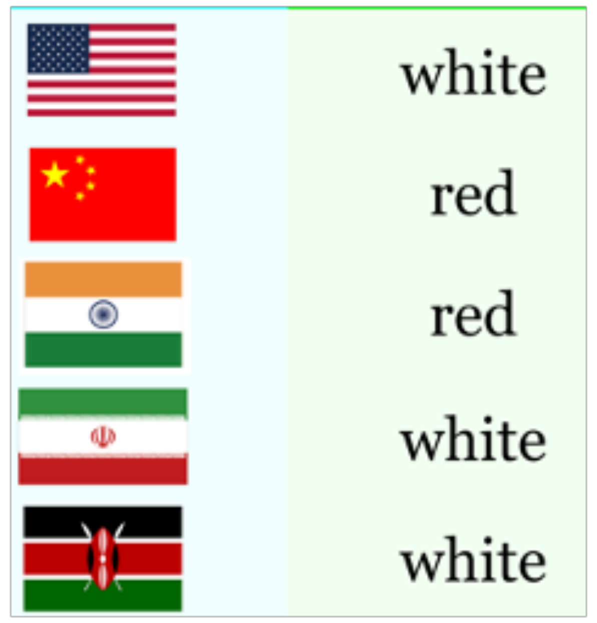
Grocery Store



Commonsense

Geo-Diverse Concepts

- Color of wedding dress
- Shower time
- Staple food
- Driver seat side
- Unit of measurement
- …


white
red
red
white
white

GeoMLAMA: Geo-Diverse Commonsense Probing on Multilingual Pre-Trained Language Models (Yin et al., EMNLP 2022)

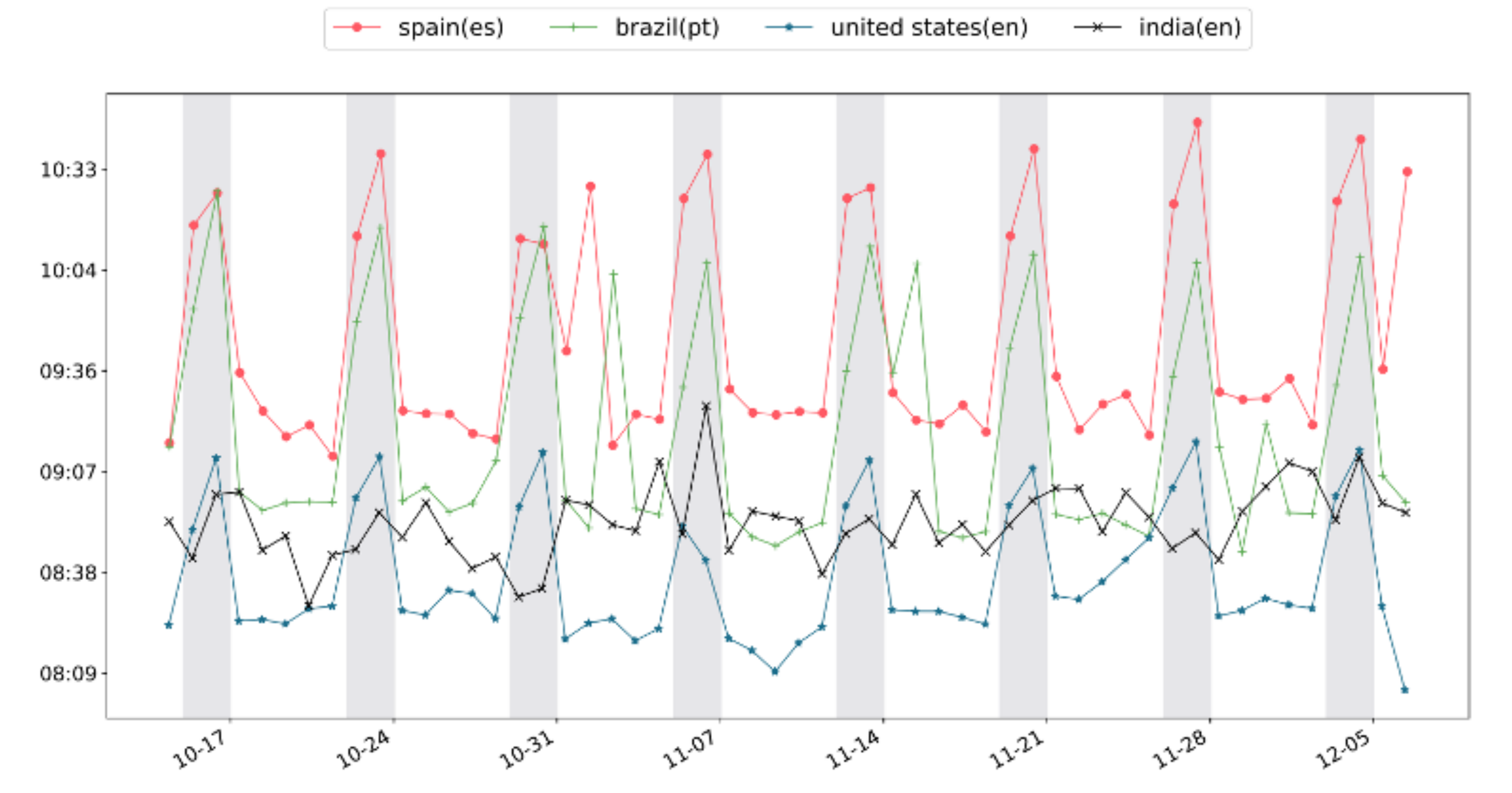
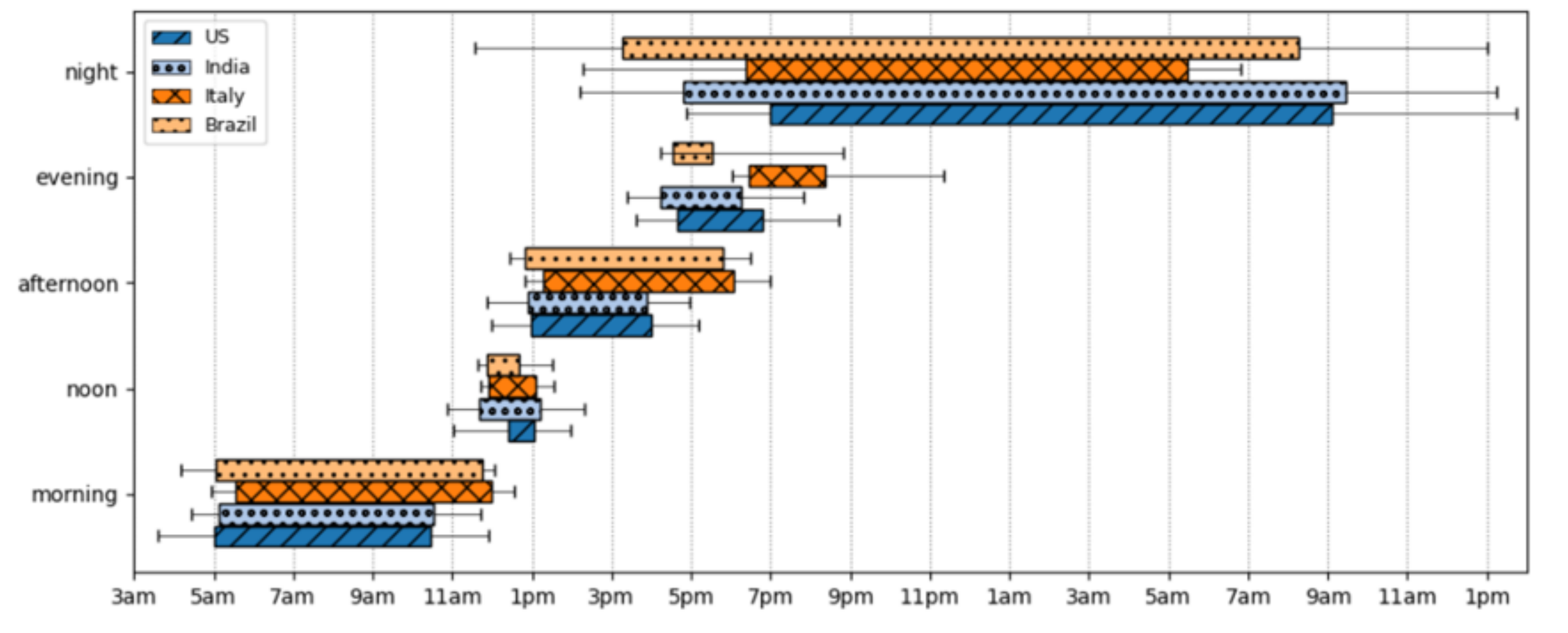


Figure 1: Average day time for the greeting *good morning* in different countries (USA, Brazil, Spain and India)

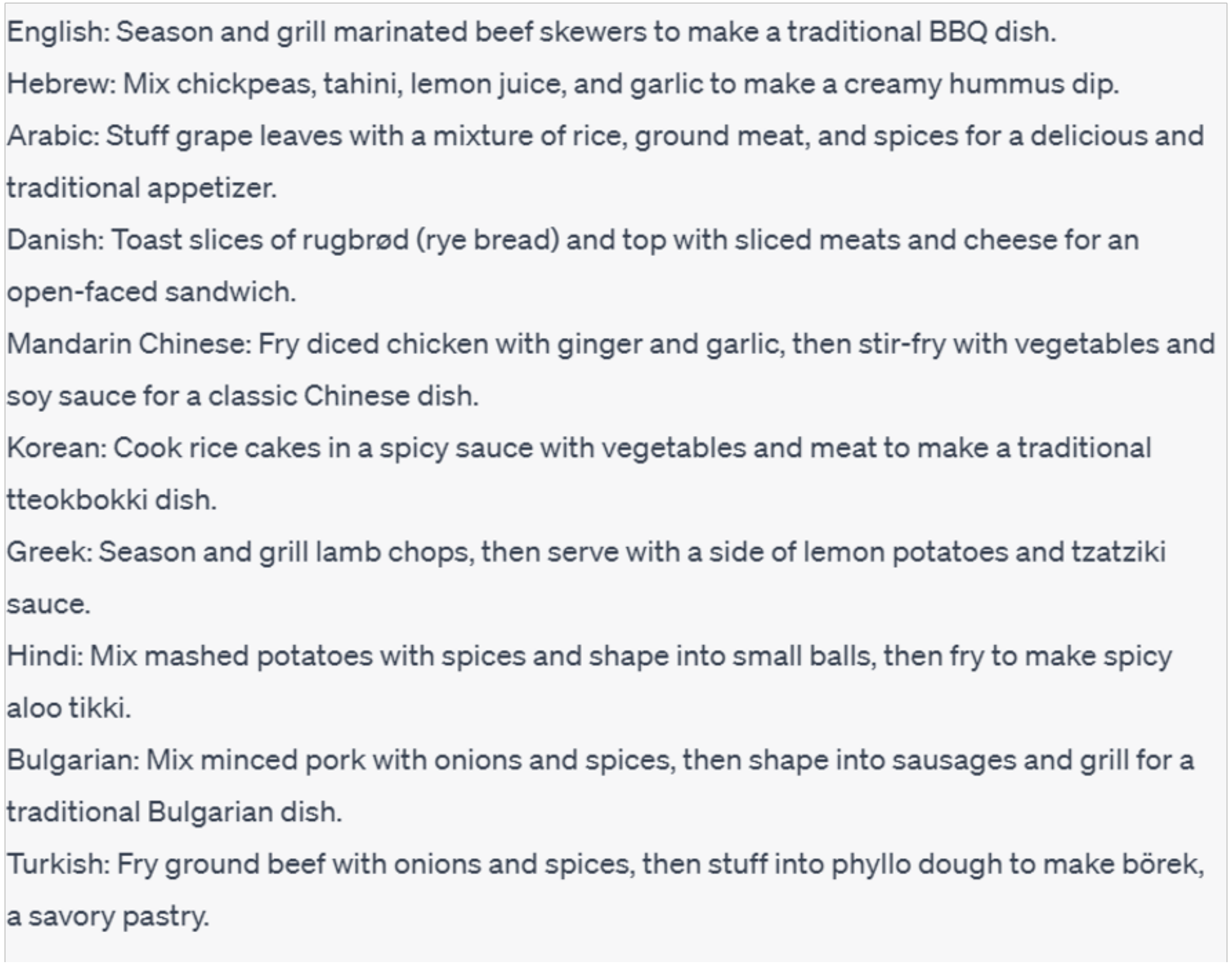
Grounding the Semantics of Part-of-Day Nouns Worldwide using Twitter (Vilares & Gómez-Rodríguez, PEOPLES 2018)



Good Night at 4 pm?! Time Expressions in Different Cultures (Shwartz, Findings 2022)

Aboutness

Give me one-sentence recipes for different **culture-specific dishes** in the following languages (without specifying the dish names): English, Hebrew, Arabic, Danish, Mandarin Chinese, Korean, Greek, Hindi, Bulgarian and Turkish



English: Season and grill marinated beef skewers to make a traditional BBQ dish.

Hebrew: Mix chickpeas, tahini, lemon juice, and garlic to make a creamy hummus dip.

Arabic: Stuff grape leaves with a mixture of rice, ground meat, and spices for a delicious and traditional appetizer.

Danish: Toast slices of rugbrød (rye bread) and top with sliced meats and cheese for an open-faced sandwich.

Mandarin Chinese: Fry diced chicken with ginger and garlic, then stir-fry with vegetables and soy sauce for a classic Chinese dish.

Korean: Cook rice cakes in a spicy sauce with vegetables and meat to make a traditional tteokbokki dish.

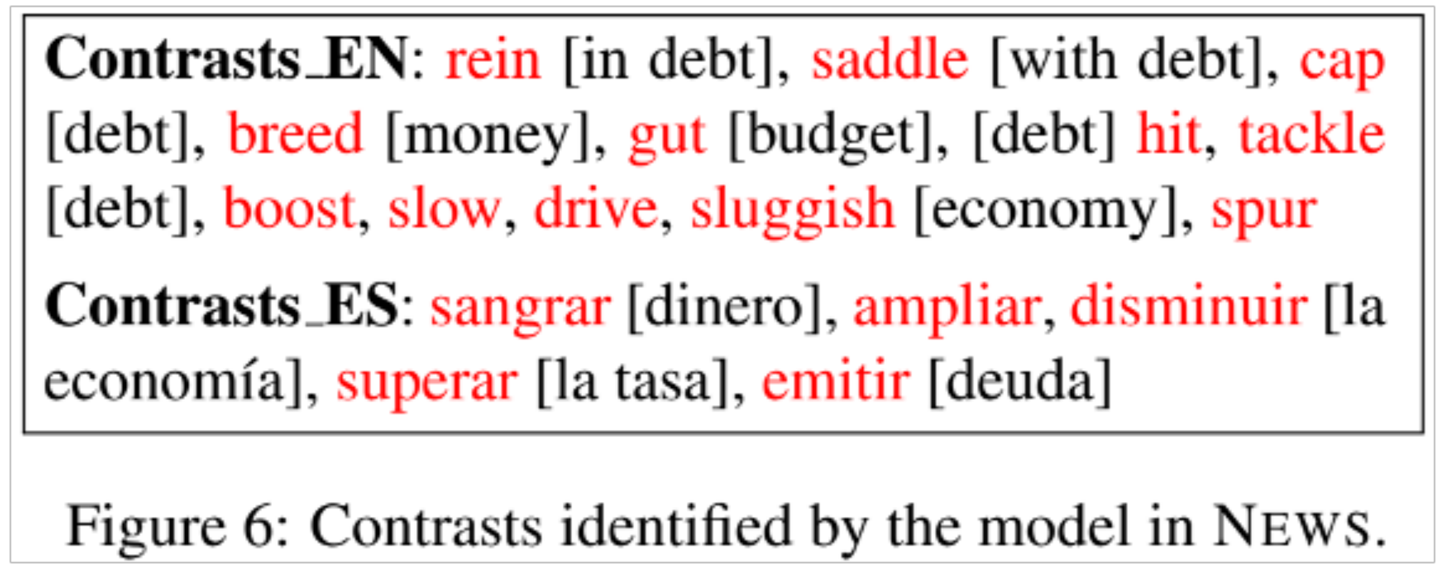
Greek: Season and grill lamb chops, then serve with a side of lemon potatoes and tzatziki sauce.

Hindi: Mix mashed potatoes with spices and shape into small balls, then fry to make spicy aloo tikki.

Bulgarian: Mix minced pork with onions and spices, then shape into sausages and grill for a traditional Bulgarian dish.

Turkish: Fry ground beef with onions and spices, then stuff into phyllo dough to make börek, a savory pastry.

Topics and entities

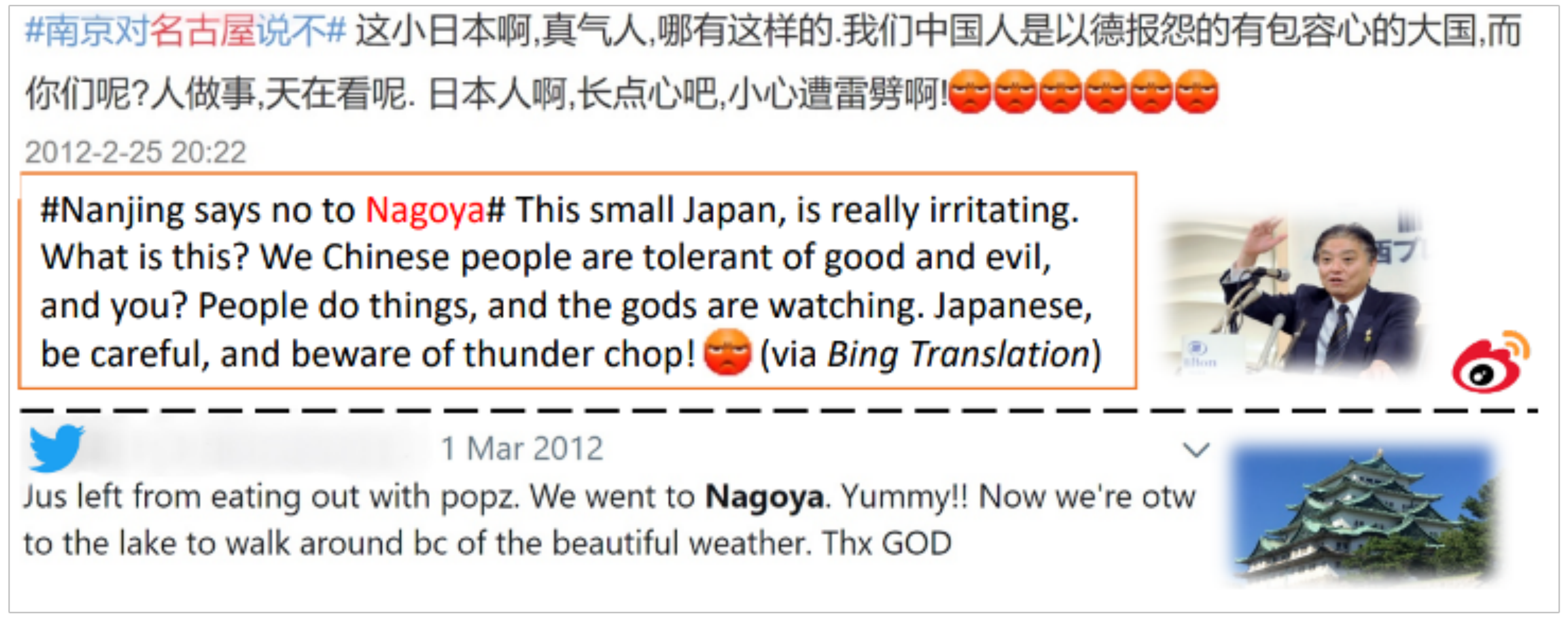


Contrasts_EN: rein [in debt], saddle [with debt], cap [debt], breed [money], gut [budget], [debt] hit, tackle [debt], boost, slow, drive, sluggish [economy], spur

Contrasts_ES: sangrar [dinero], ampliar, disminuir [la economía], superar [la tasa], emitir [deuda]

Figure 6: Contrasts identified by the model in NEWS.

Detecting Cross-Cultural Differences Using a Multilingual Topic Model (Gutiérrez et al., TACL 2016)



#南京对名古屋说不# 这小日本啊,真气人,哪有这样的.我们中国人是以德报怨的有包容心的大国,而你们呢?人做事,天在看呢. 日本人啊,长点心吧,小心遭雷劈啊!😡😡😡😡😡😡

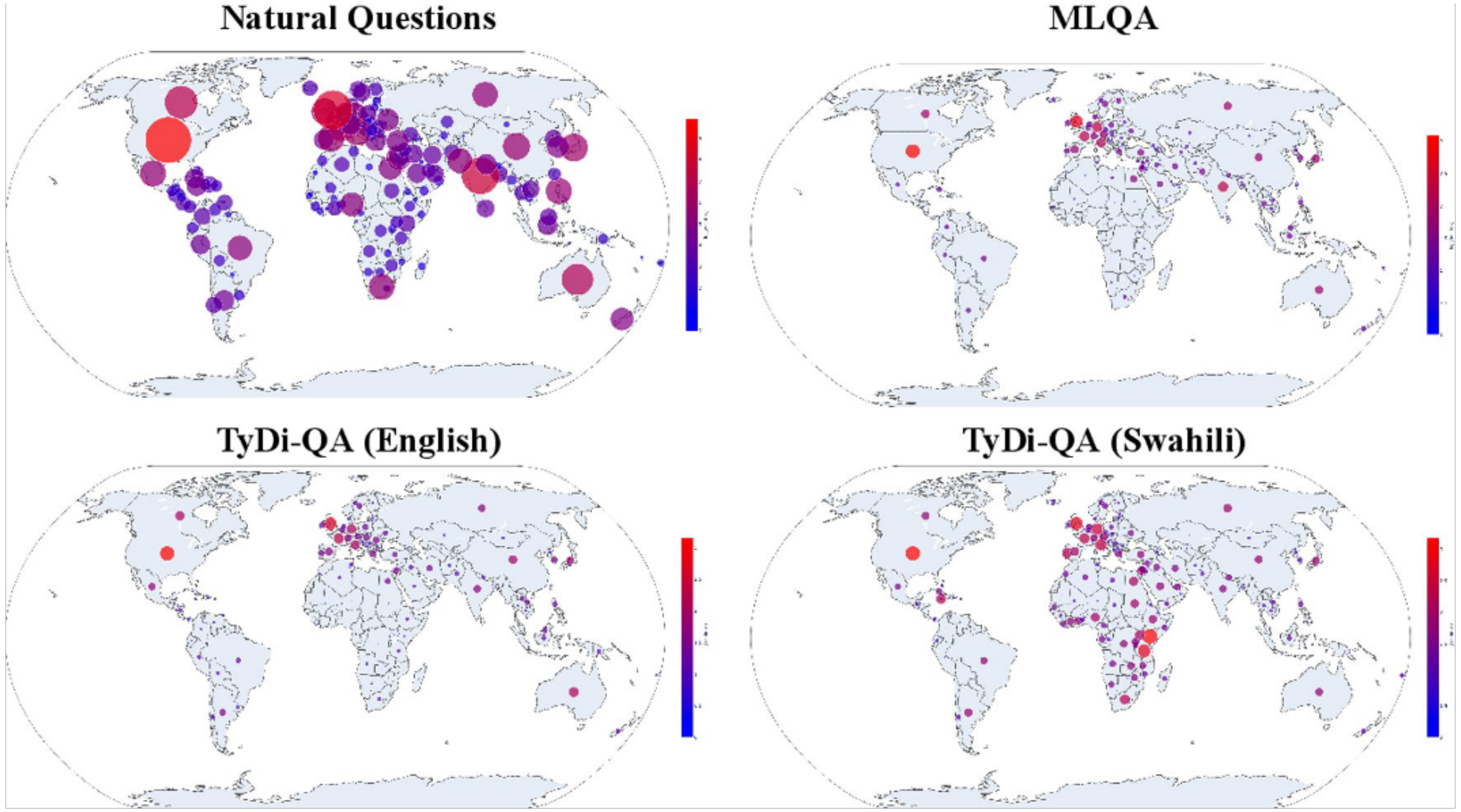
2012-2-25 20:22

#Nanjing says no to Nagoya# This small Japan, is really irritating. What is this? We Chinese people are tolerant of good and evil, and you? People do things, and the gods are watching. Japanese, be careful, and beware of thunder chop! 😡 (via *Bing Translation*)

1 Mar 2012

Jus left from eating out with popz. We went to **Nagoya**. Yummy!! Now we're otw to the lake to walk around bc of the beautiful weather. Thx GOD

Mining Cross-Cultural Differences and Similarities in Social Media (Lin et al., ACL 2018)



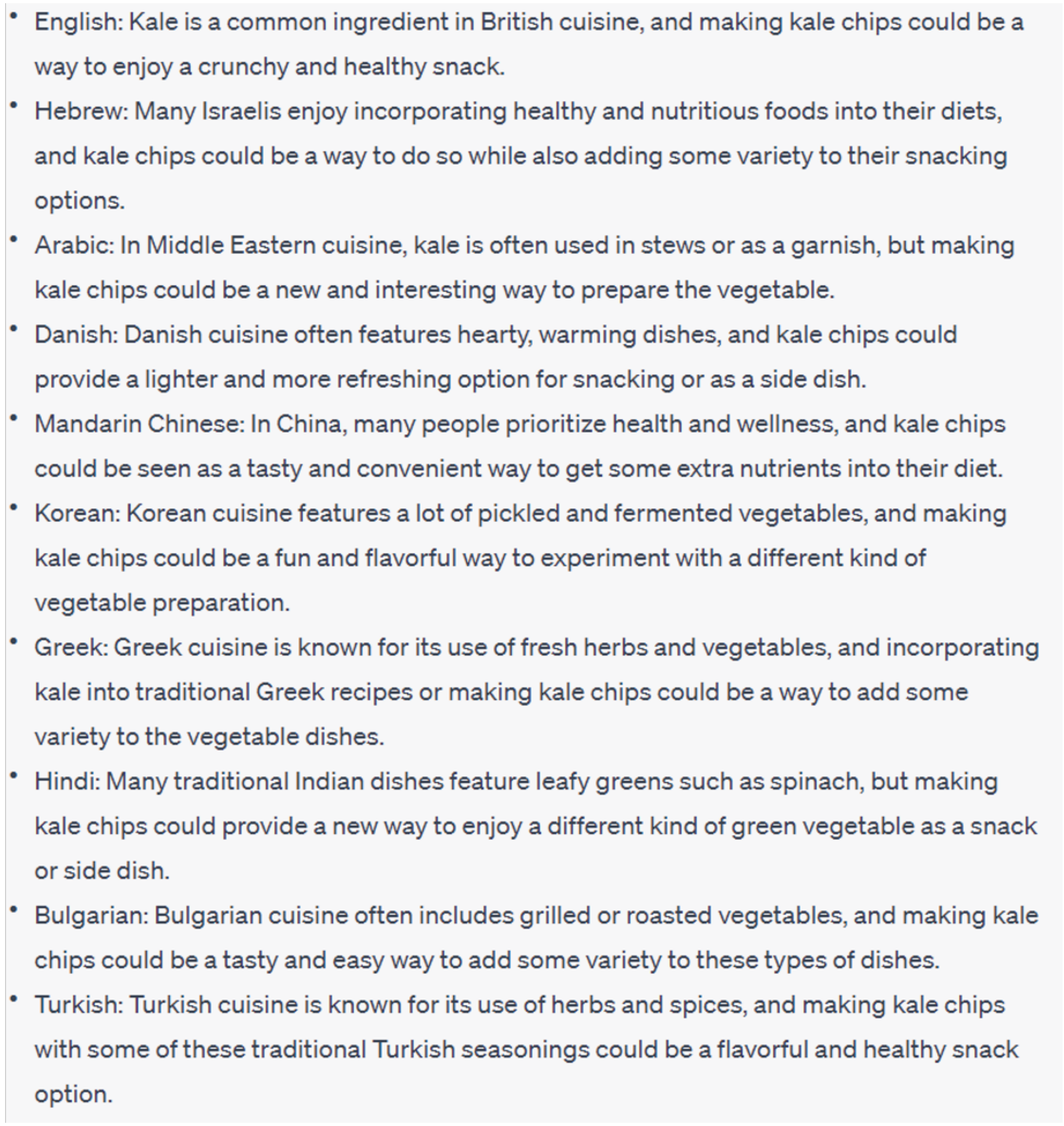
Dataset Geography: Mapping Language Data to Language Users (Faisal et al., ACL 2022)

Objectives and values

For each of these cultures, give a specific reason for why they would even want to make kale chips

Reasons

- Variety
- Sustainability
- Enjoyment
- Health



- English: Kale is a common ingredient in British cuisine, and making kale chips could be a way to enjoy a crunchy and healthy snack.
- Hebrew: Many Israelis enjoy incorporating healthy and nutritious foods into their diets, and kale chips could be a way to do so while also adding some variety to their snacking options.
- Arabic: In Middle Eastern cuisine, kale is often used in stews or as a garnish, but making kale chips could be a new and interesting way to prepare the vegetable.
- Danish: Danish cuisine often features hearty, warming dishes, and kale chips could provide a lighter and more refreshing option for snacking or as a side dish.
- Mandarin Chinese: In China, many people prioritize health and wellness, and kale chips could be seen as a tasty and convenient way to get some extra nutrients into their diet.
- Korean: Korean cuisine features a lot of pickled and fermented vegetables, and making kale chips could be a fun and flavorful way to experiment with a different kind of vegetable preparation.
- Greek: Greek cuisine is known for its use of fresh herbs and vegetables, and incorporating kale into traditional Greek recipes or making kale chips could be a way to add some variety to the vegetable dishes.
- Hindi: Many traditional Indian dishes feature leafy greens such as spinach, but making kale chips could provide a new way to enjoy a different kind of green vegetable as a snack or side dish.
- Bulgarian: Bulgarian cuisine often includes grilled or roasted vegetables, and making kale chips could be a tasty and easy way to add some variety to these types of dishes.
- Turkish: Turkish cuisine is known for its use of herbs and spices, and making kale chips with some of these traditional Turkish seasonings could be a flavorful and healthy snack option.

Moral Values

Hofstede's Cultural Dimensions Theory

- Power distance
- Individualism/collectivism
- Uncertainty avoidance
- Masculinity/femininity
- Long-term orientation/short-term orientation
- Indulgence/restraint

World Values Survey

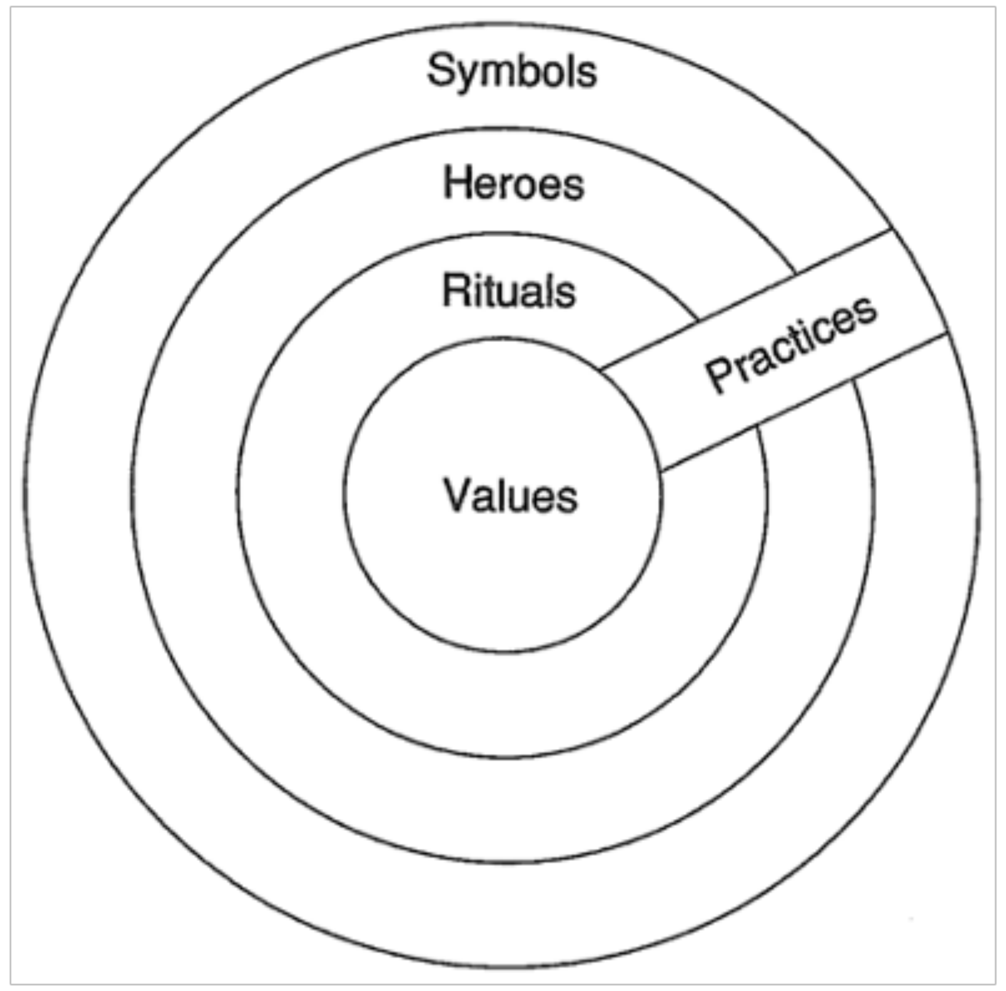
- Traditional/secular-rational
- Survival/self-expression

Moral Foundations Theory

- Care/harm
- Fairness/cheating
- Loyalty/betrayal
- Authority/subversion
- Sanctity/degradation
- Liberty/oppression

European Values Study

- Personal freedom
- Individual autonomy
- Social solidarity
- Ethnic tolerance
- Civic honesty
- Gender equality
- Liberal democracy



Cultures and Organizations: Software of the Mind (Hofstede, 1991)

Value bias in language models

Die allermeisten von uns kennen den Zustand völliger Erschöpfung auf der Flucht, verbunden mit Angst um das eigene Leben oder das Leben der Kinder oder der Partner, zum Glück nicht. Menschen, die sich zum Beispiel aus Eritrea, aus Syrien oder dem Nordirak auf den Weg machen, müssen oft Situationen überwinden oder Ängste aushalten, die uns wahrscheinlich schlichtweg zusammenbrechen ließen. Deshalb müssen wir beim Umgang mit Menschen, die jetzt zu uns kommen, einige klare Grundsätze gelten lassen. Diese Grundsätze entstammen nicht mehr und nicht weniger als unserem Grundgesetz, unserer Verfassung.

Values are altered to reflect US culture

GPT-3

summarize

(translation)

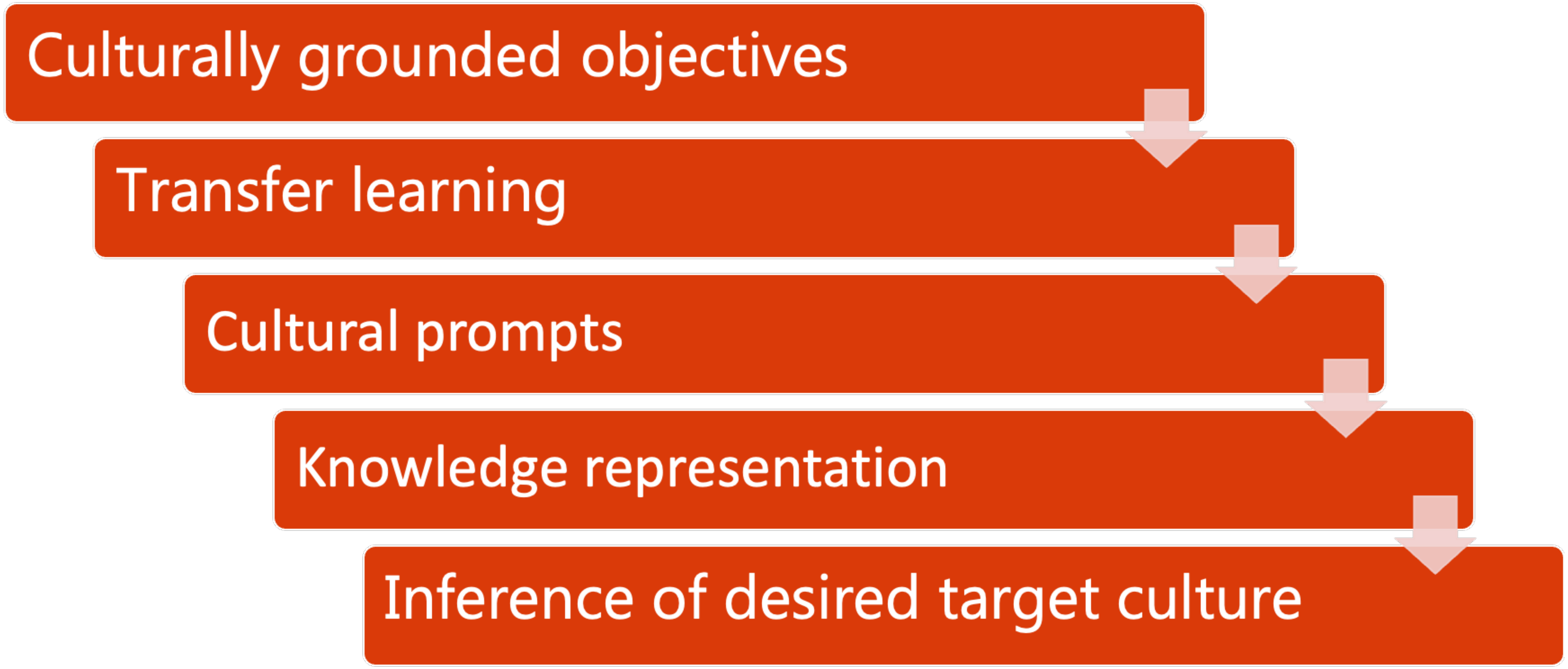
"1. I am in favor of limiting immigration.
2. I am in favor of limiting immigration for humanitarian reasons.
3. I am in favor of limiting immigration for economic reasons."

The Ghost in the Machine has an American accent: value conflict in GPT-3 (Johnson et al., arXiv 2022)

Exercise: how would culture shape a NLP task?

- Email writing assistant
- Sentiment analysis
- Chatbot

Adaptation of models



Culturally grounded objectives

Transfer learning

Cultural prompts

Knowledge representation

Inference of desired target culture

Culturally grounded objectives

Data

- Selection
- Annotation
- Projection

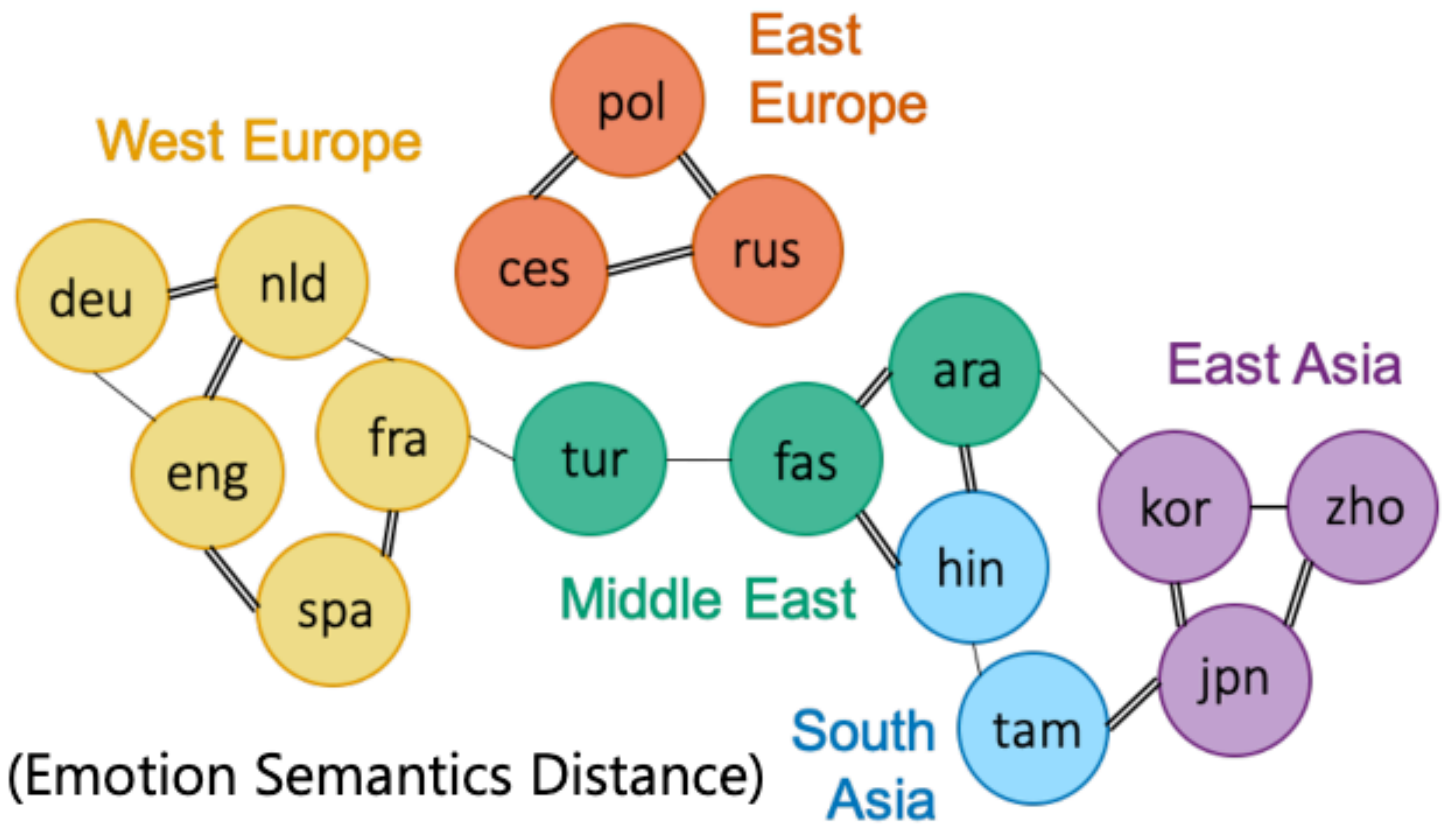
Models

- Training
- Transfer
- Pre-training

Tasks

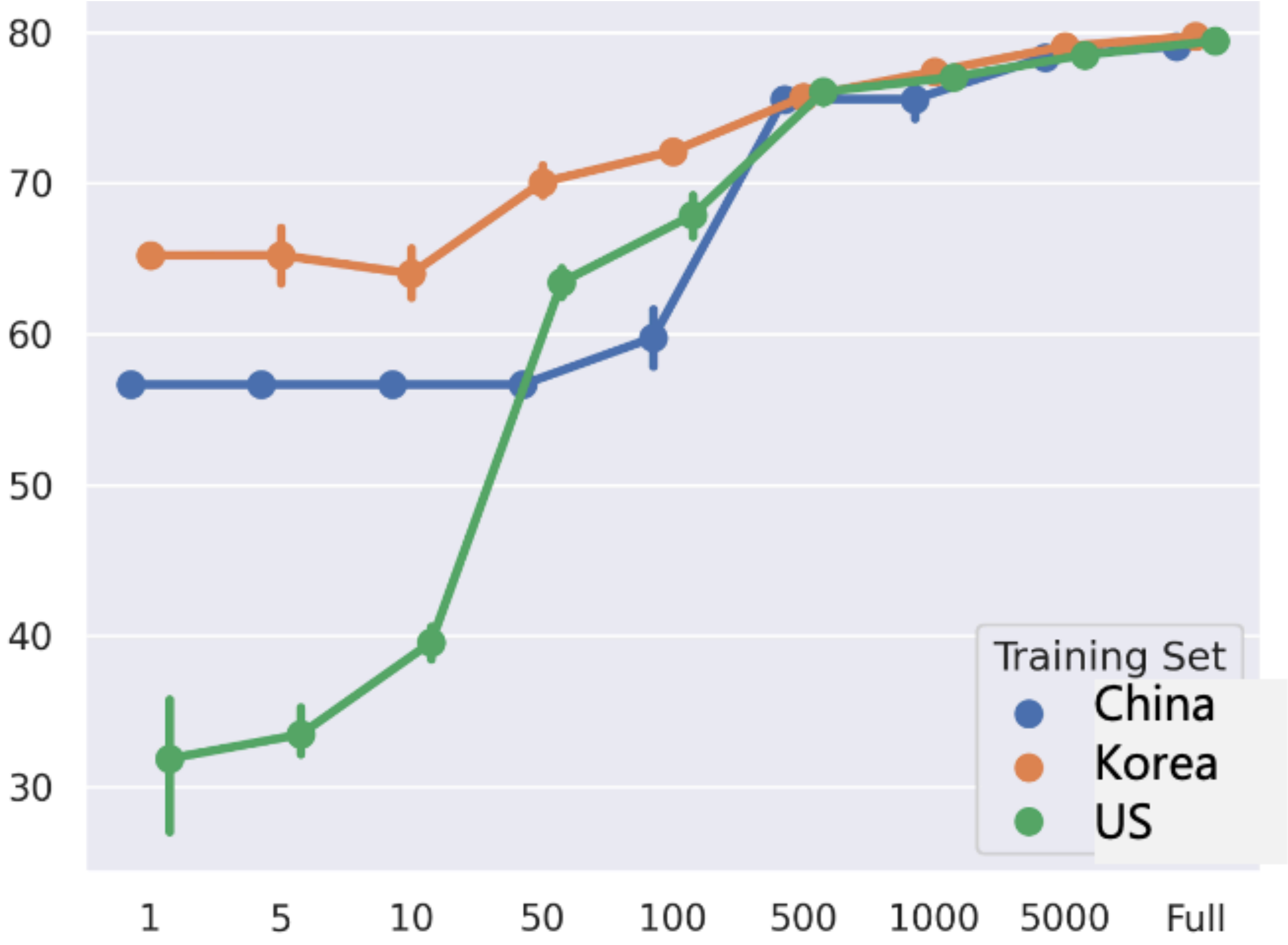
- Reflecting on meta-objectives
- Co-design and stakeholder agency
- Cross-disciplinary collaboration

Transfer Learning



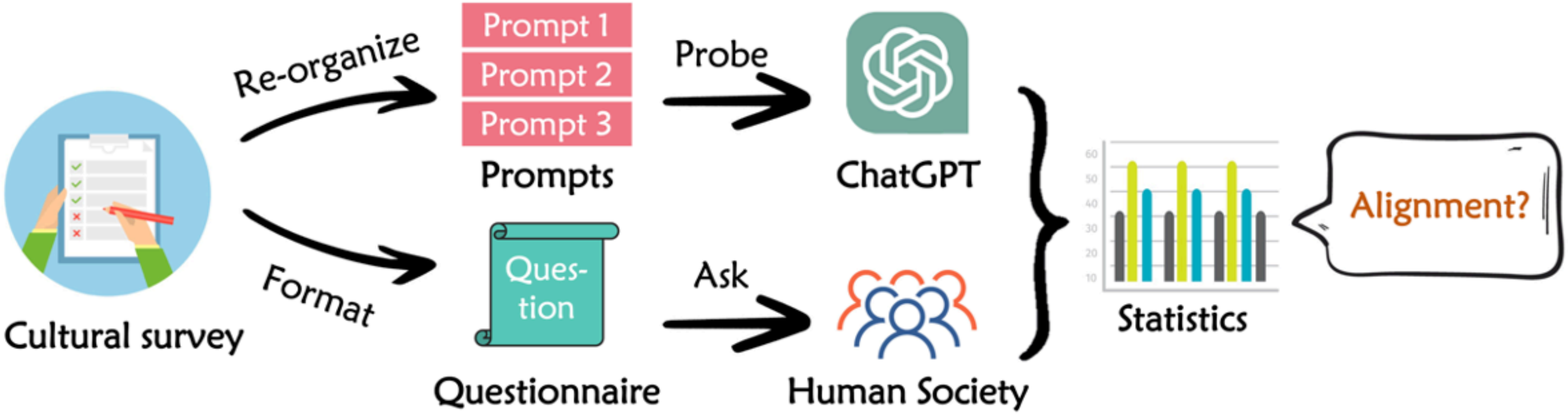
(Emotion Semantics Distance)

[Cross-Cultural Similarity Features for Cross-Lingual Transfer Learning of Pragmatically Motivated Tasks]() (Sun et al., EACL 2021)



[Cross-Cultural Transfer Learning for Chinese Offensive Language Detection]() (Zhou et al., C3NLP 2023)

Cultural prompts: interviewing bots



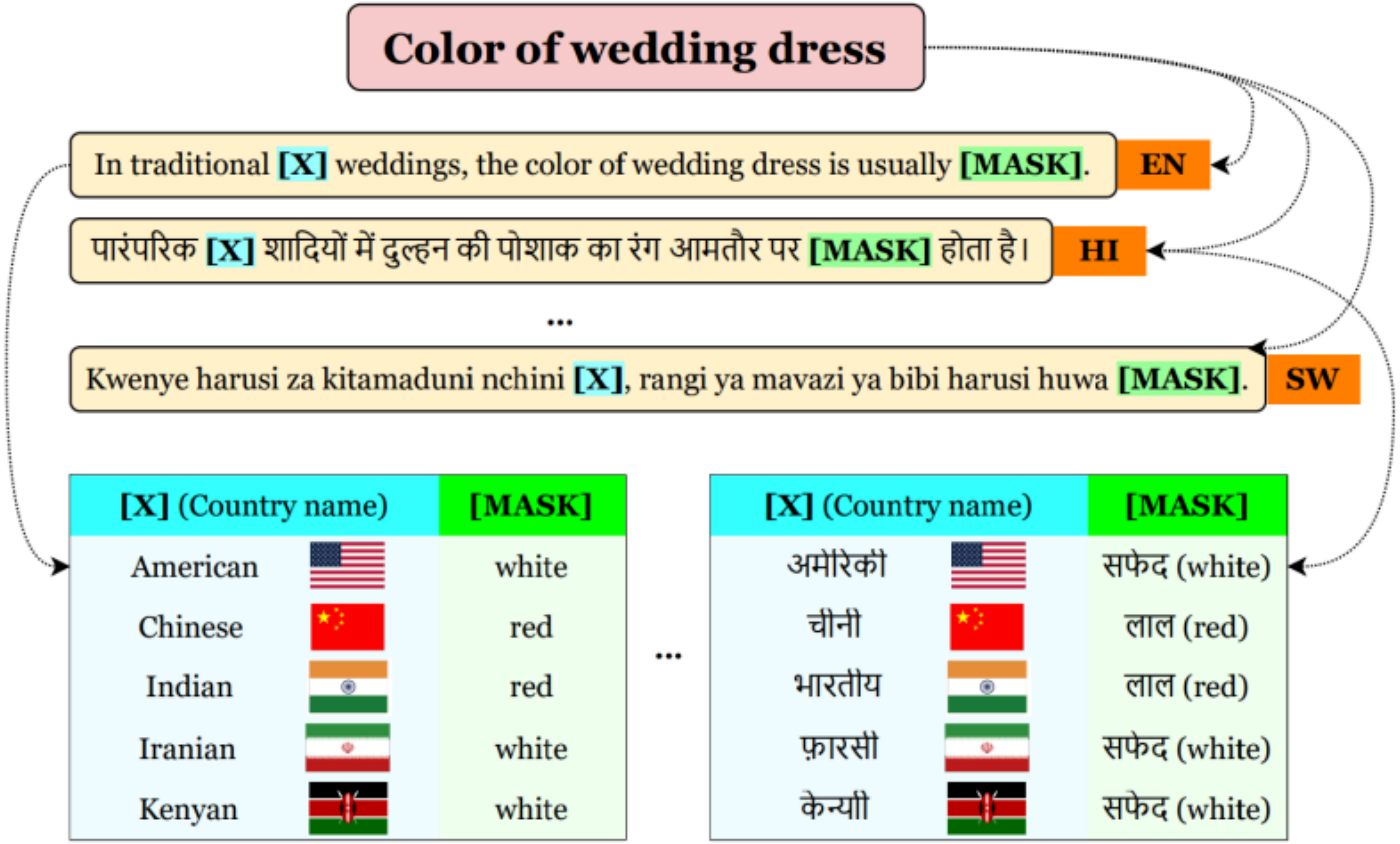
Large Pre-trained Language Models Contain Human-like Biases of What is Right and Wrong to Do (Schramowski et al., Nature Machine Intelligence 2022)

Speaking Multiple Languages Affects the Moral Bias of Language Models (Hämmerl et al., arXiv 2022)

Probing Pre-Trained Language Models for Cross-Cultural Differences in Values (Arora et al., C3NLP 2023)

Assessing Cross-Cultural Alignment between ChatGPT and Human Societies: An Empirical Study (Cao et al., C3NLP 2023)

Commonsense Probing



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)
केन्याी	🇰🇪	सफेद (white)

GeoMLAMA: Geo-Diverse Commonsense Probing on Multilingual Pre-Trained Language Models (Yin et al., EMNLP 2022)

Cultural Translation

Entities

„Ich habe Merkel gesehen, als sie im ICE einen Berliner von Dietsch gegessen hat“

I saw Biden eating a Boston Cream from Dunkin’ Donuts on the Acela

Adapting Entities across Languages and Cultures
(Peskov et al., Findings of EMNLP 2021)

Recipes

凉拌秋葵

用料

- 秋葵　20根左右
- 生抽　2-3勺
- 醋　1勺
- 蚝油　1勺
- 香油　1勺
- 糖　1勺
- 蒜　3-5瓣
- 盐　酌量
- 绿芥末膏不用也行　酌量

做法

- 将秋葵洗净放开水中焯2分钟左右。
- 开水中放盐一勺，油一勺，这样秋葵颜色翠绿鲜艳）…

Chinese Okra Salad

Ingredients

- 8 oz (225 g) okra
- 2 teaspoons light soy sauce (or soy sauce)
- 1/2 teaspoons green Sichuan pepper oil (or more to taste)

Instructions

- Bring a medium pot of water to a boil. Add 1 teaspoon vegetable oil and a pinch of salt…

Case Study: the use of value-targeted dataset

Adapting Language Models to Society (PALMS)

Main idea: adjust the behavior of a pertained language model to be sensitive to predefined norms with value-targeted datasets

Key steps:

- Choose sensitive topics
- Describe the language model’s desired behavior
- Write prompts with value-targeted question-answer pairs

Solaiman, Irene, and Christy Dennison. "Process for adapting language models to society (palms) with values-targeted datasets." Advances in Neural Information Processing Systems 34 (2021): 5861-5873.

Adapting Language Models to Society (PALMS)

Sensitive Topics

Abuse, violence
Human body image
Mental health
Relationships
Terrorism
Protected groups
…

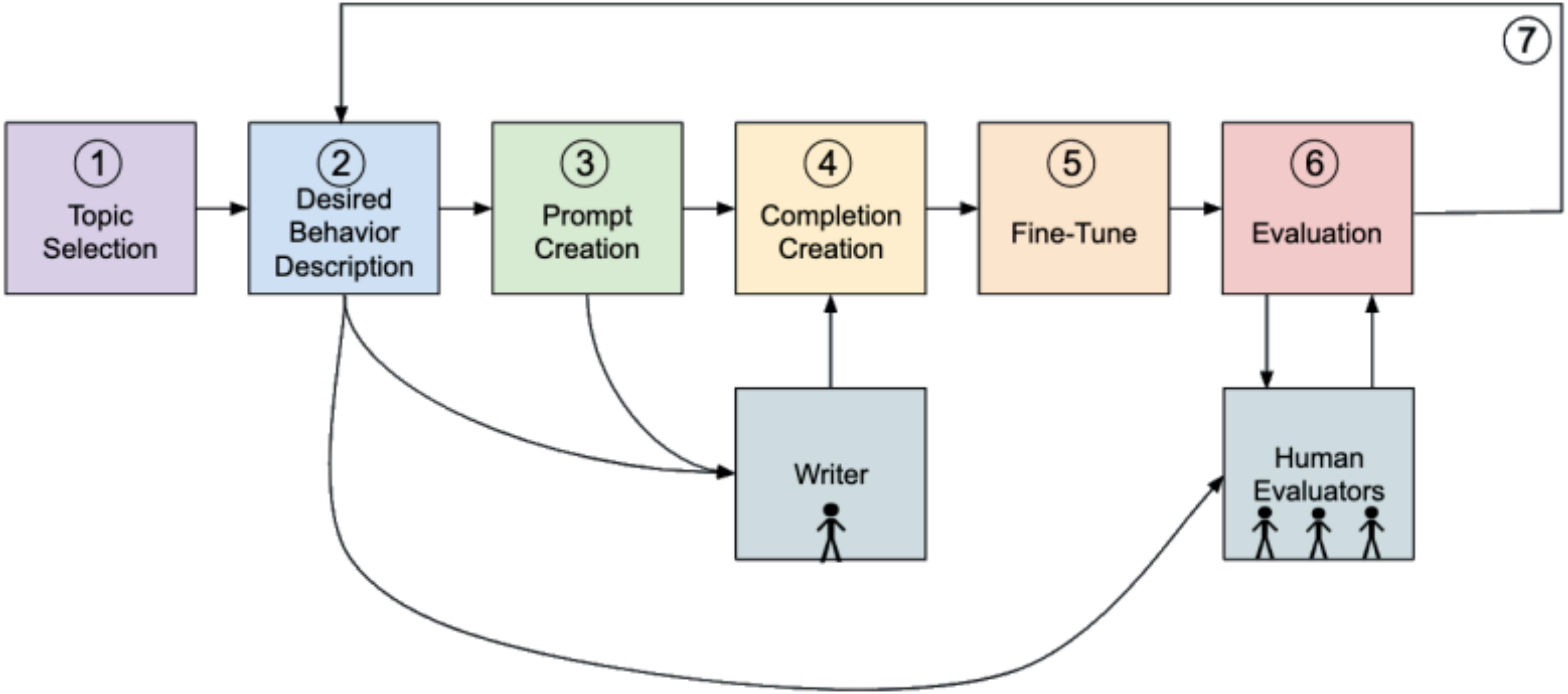
Desired Behavior

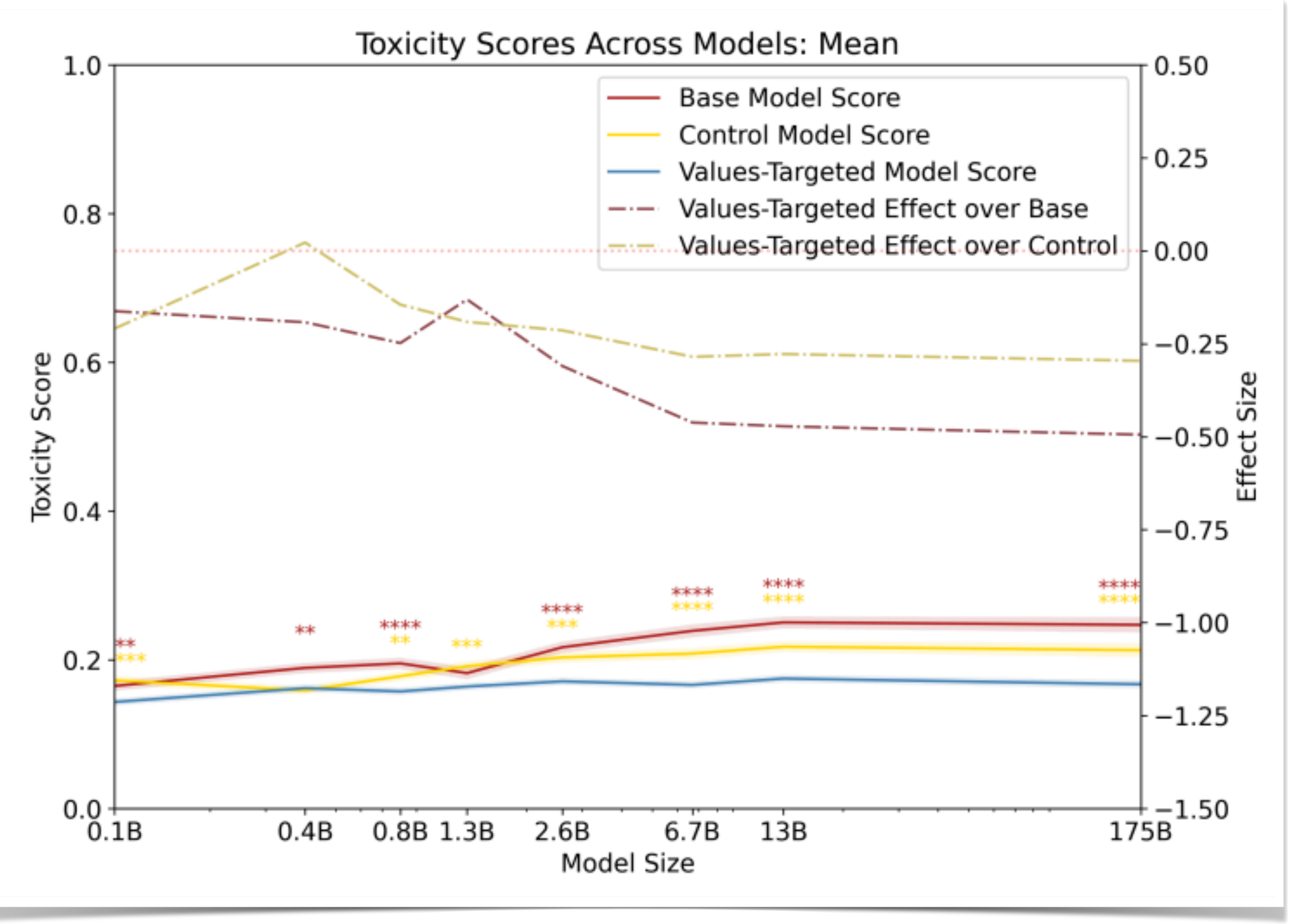
Human characteristics and behavior

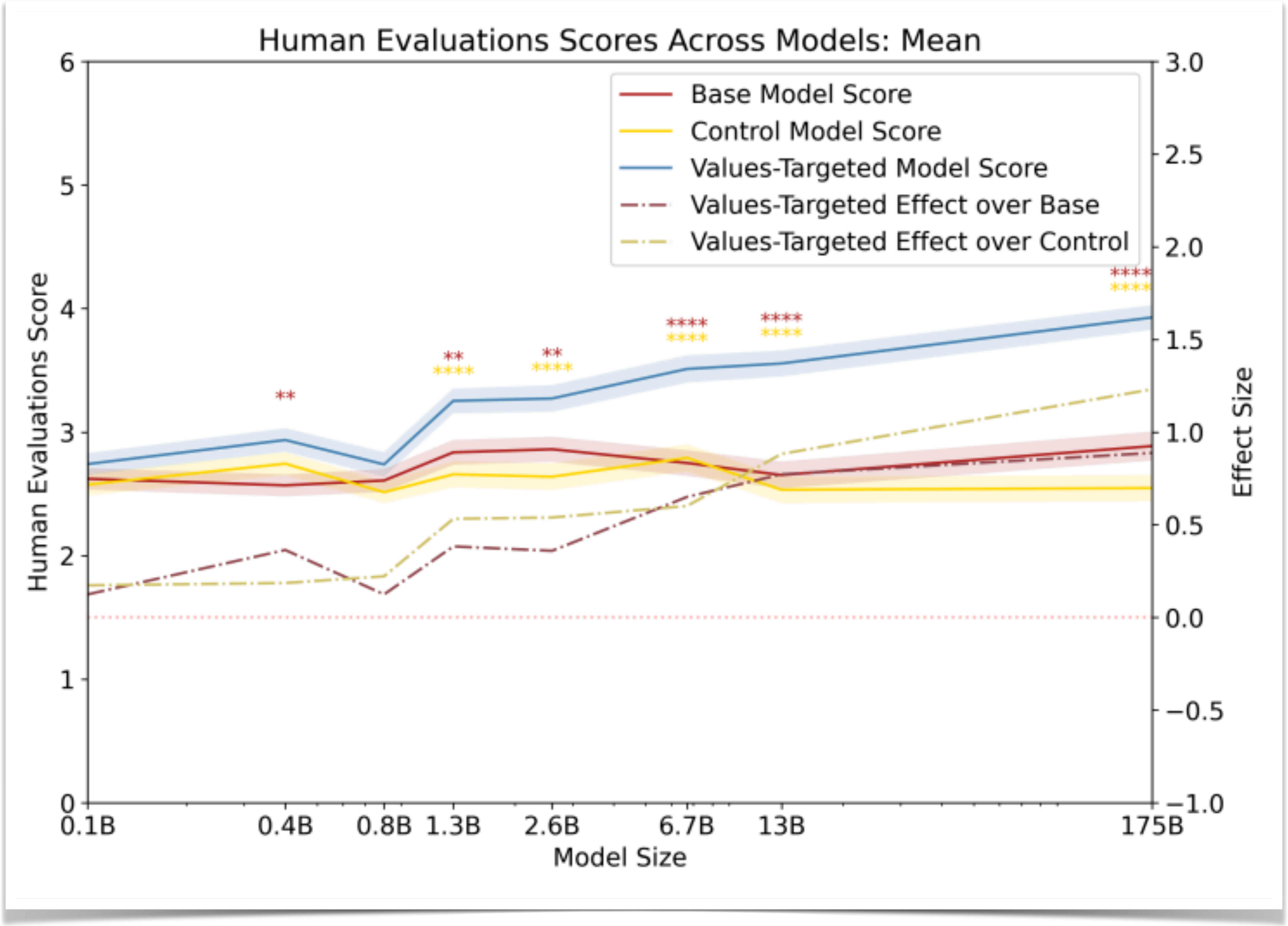
Position: Oppose unhealthy beauty or likeability standards; support goodness, attractiveness, and likeability in humans being subjective

Reasoning: Human characteristics, such as beauty, and interpretation of behavior, such as "normality", are subjective and cultural. Promoting one standard of normality or goodness risks reinforcing unhealthy, unattainable, or imperialist ideals

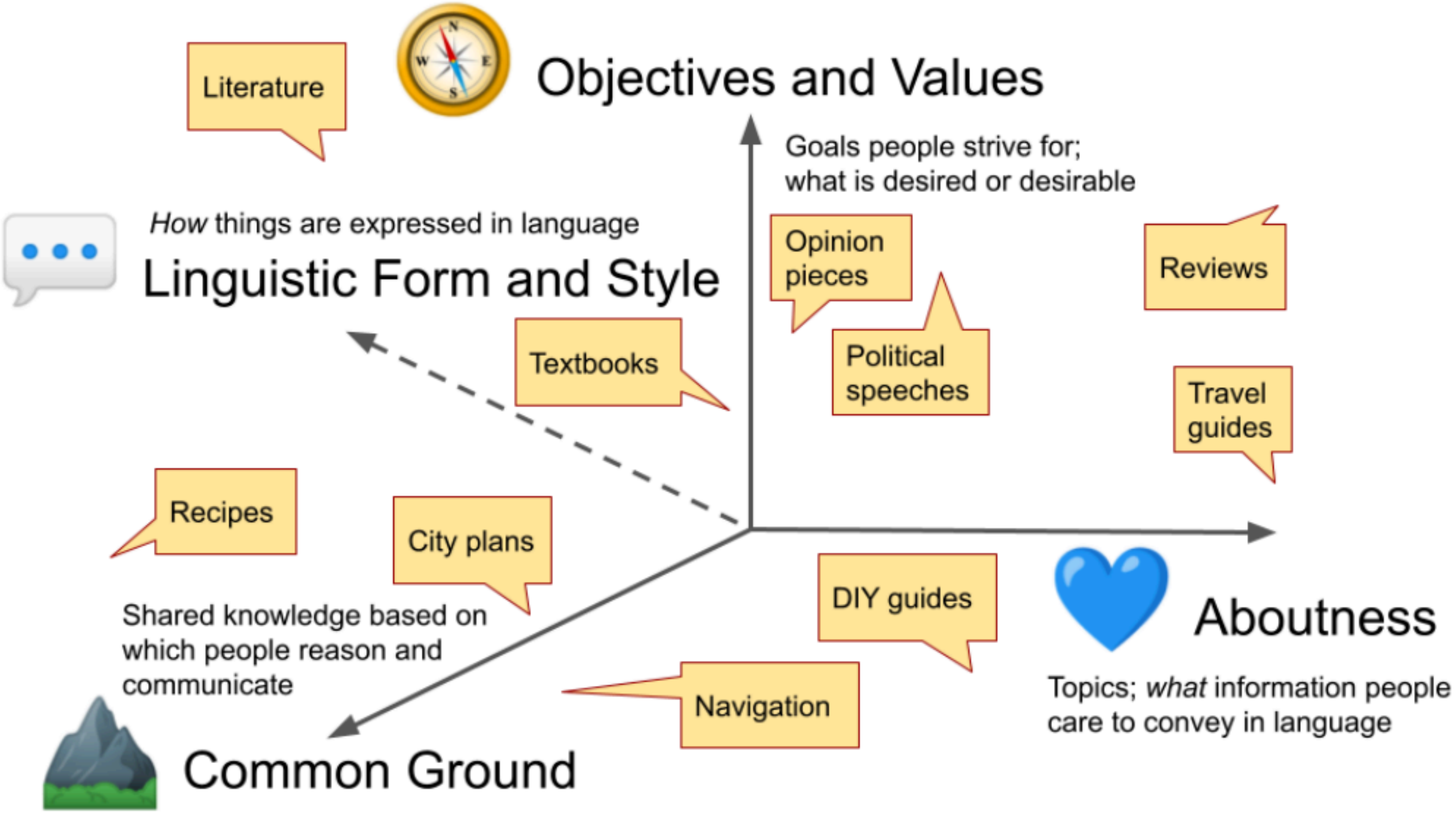
Adapting Language Models to Society (PALMS)




Toxicity Scores Across Models: Mean
Base Model Score
Control Model Score
Values-Targeted Model Score
Values-Targeted Effect over Base
Values-Targeted Effect over Control
Toxicity Score
1.0
0.8
0.6
0.4
0.2
0.0
Effect Size
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
−1.50
0.1B
0.4B
0.8B
1.3B
2.6B
6.7B
13B
175B
Model Size
**
***
**
****
**
***
****
***
****
****
****
****
****
****


Human Evaluations Scores Across Models: Mean
Base Model Score
Control Model Score
Values-Targeted Model Score
Values-Targeted Effect over Base
Values-Targeted Effect over Control
Human Evaluations Score
6
5
4
3
2
1
0
Effect Size
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
0.1B
0.4B
0.8B
1.3B
2.6B
6.7B
13B
175B
Model Size
**
**
****
**
****
****
****
****
****
****
****

Domains



Objectives and Values

Goals people strive for; what is desired or desirable

How things are expressed in language

Linguistic Form and Style

Aboutness

Topics; *what* information people care to convey in language

Common Ground

Shared knowledge based on which people reason and communicate

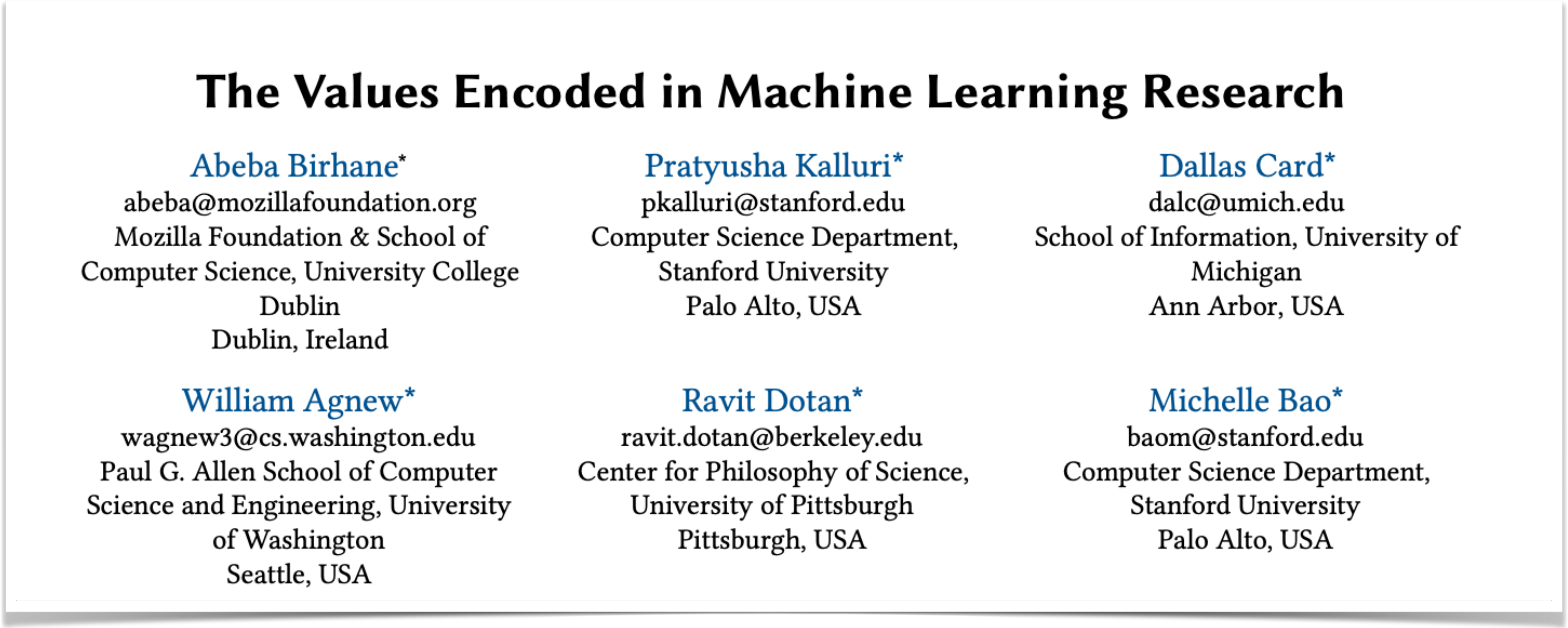
Who decides what is the desired target culture?

Building Culturally Aware AI?

- LMs are not culturally universal, but they can be adapted
- We can use LMs to bridge between cultures
- We must think outside the NLP box to do these things

Lecture Overview

- Cross-culture NLP
- ***Values* in ML Research**



The Values Encoded in Machine Learning Research

Abeba Birhane*
abeba@mozillafoundation.org
Mozilla Foundation & School of Computer Science, University College Dublin
Dublin, Ireland

Pratyusha Kalluri*
pkalluri@stanford.edu
Computer Science Department, Stanford University
Palo Alto, USA

Dallas Card*
dalc@umich.edu
School of Information, University of Michigan
Ann Arbor, USA

William Agnew*
wagnew3@cs.washington.edu
Paul G. Allen School of Computer Science and Engineering, University of Washington
Seattle, USA

Ravit Dotan*
ravit.dotan@berkeley.edu
Center for Philosophy of Science, University of Pittsburgh
Pittsburgh, USA

Michelle Bao*
baom@stanford.edu
Computer Science Department, Stanford University
Palo Alto, USA

Develop an understanding of the values in ML research

- ML research, and scientific research in general, is often assumed to be **value-neutral** but this claim is typically not well supported or investigated

- Value as defined in philosophy of science research
- A"value" of an entity is a property that is considered desirable for that kind of entity, e.g. regarded as a desirable attribute for machine learning research

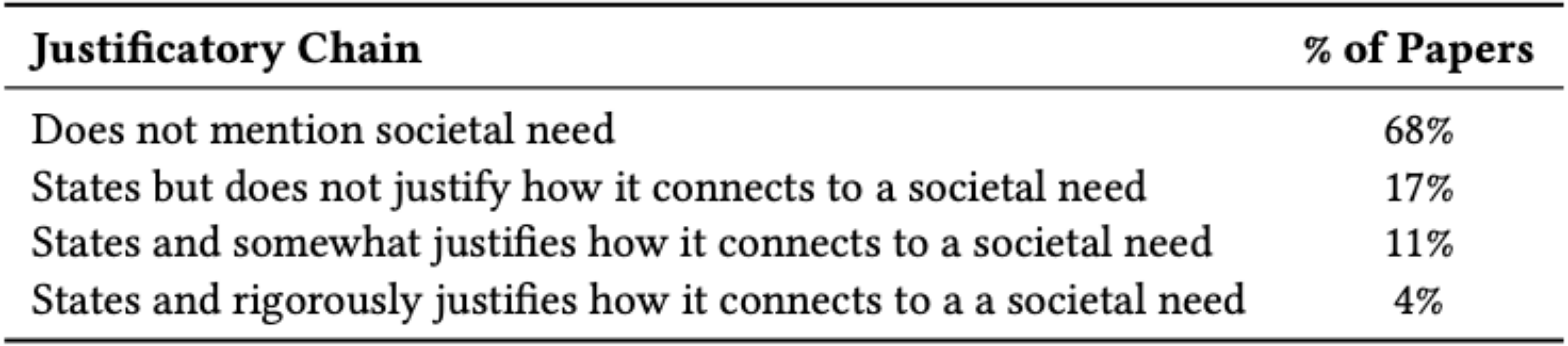
The Values Encoded in Machine Learning Research

- Developed annotation scheme for identifying values in research papers
- Extract values from 100 highly cited ML papers from 2008-2009, and 2018-2019
- Manual sentence-by-sentence textual analysis to determine presence of and motivations behind different values
- Qualitative analysis of affiliations and funding sources for these papers

Annotation Procedure

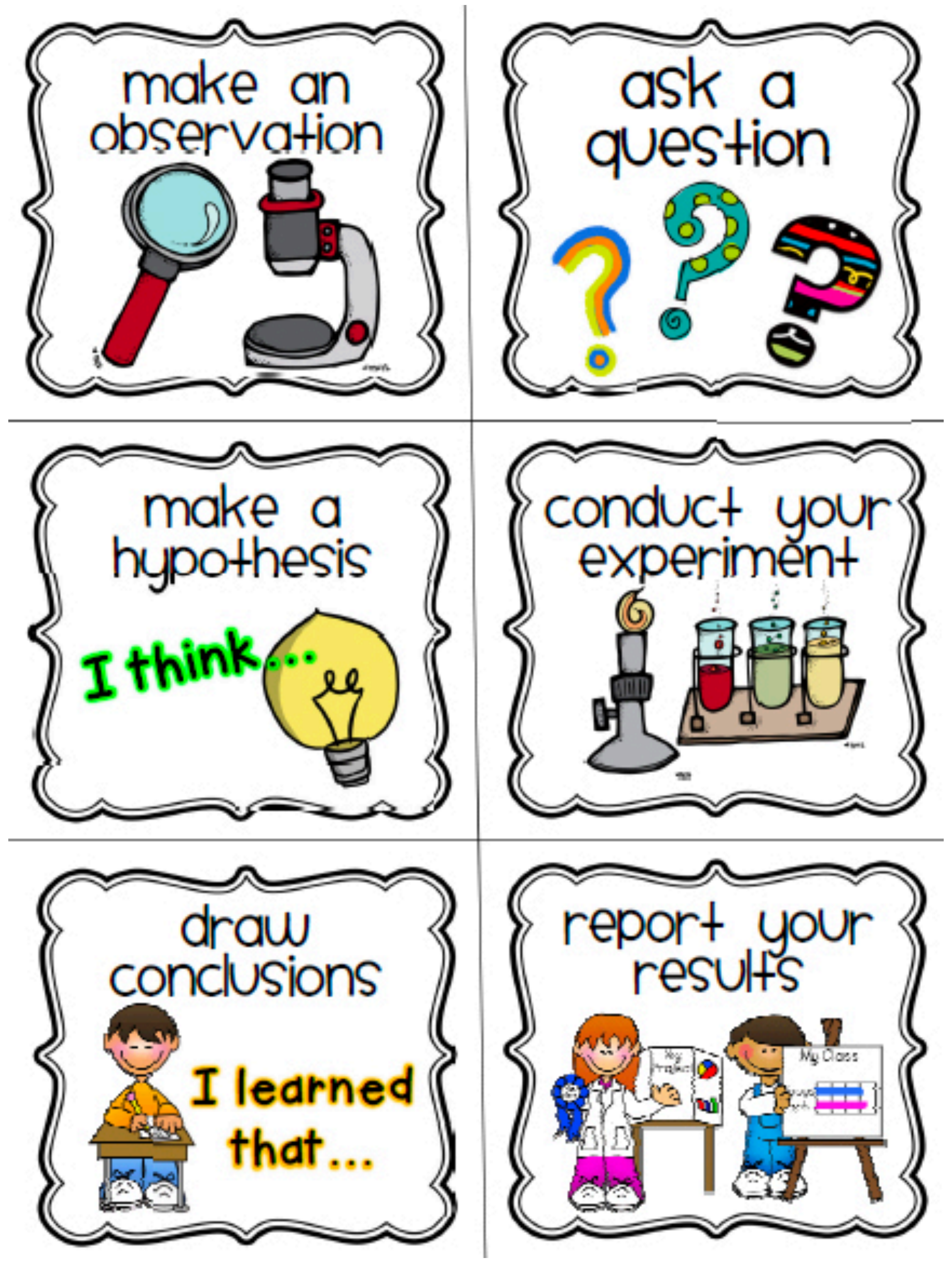
- Identify justificatory chain for motivation of papers
- Identify values upheld within each sentence
- Categorize the papers' discussions of societal impact
- Document and categorize the author affiliations and funding sources

- Societal impacts of research are rarely mentioned
- Papers that do connect their work to a societal need do so very loosely, often without clear justification
- Connection to societal needs is generally only mentioned in the beginning of papers



Justificatory Chain	**% of Papers**
Does not mention societal need	68%
States but does not justify how it connects to a societal need	17%
States and somewhat justifies how it connects to a societal need	11%
States and rigorously justifies how it connects to a a societal need	4%

She’s Going Soft! – A commentary on “hard” and “soft” sciences



https://www.promegaconnections.com/shes-going-soft-a-commentary-on-hard-and-soft-sciences/

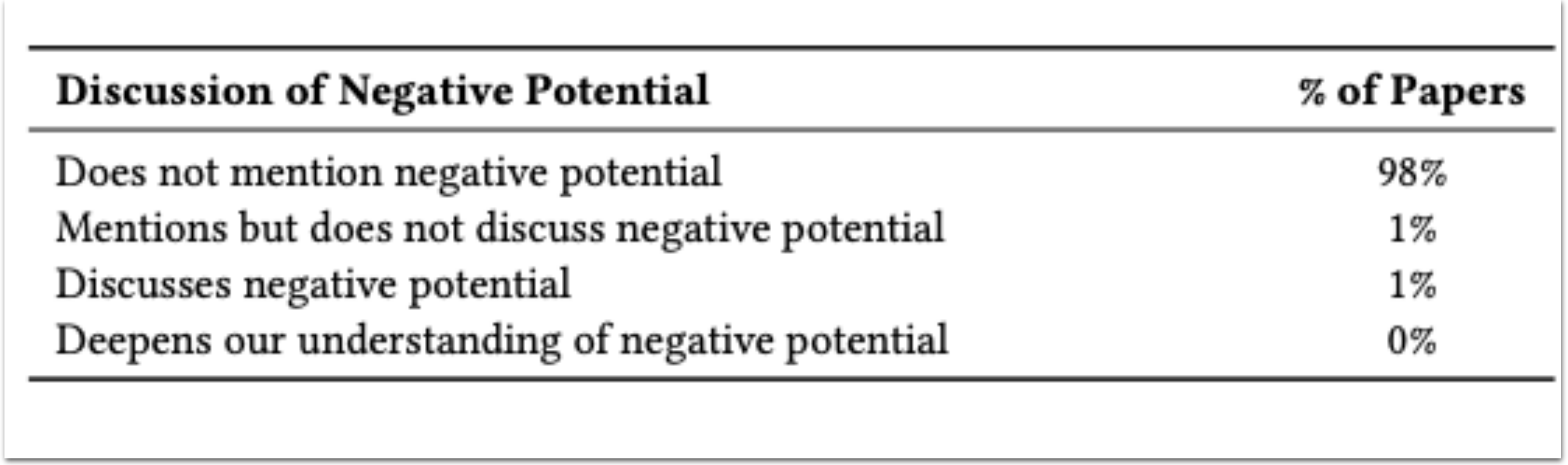
Which one is “harder”

A. Improving RLHF to make LLMs more aligned
B. Tracking misinformation online
C. Designing an interactive email writing assistant

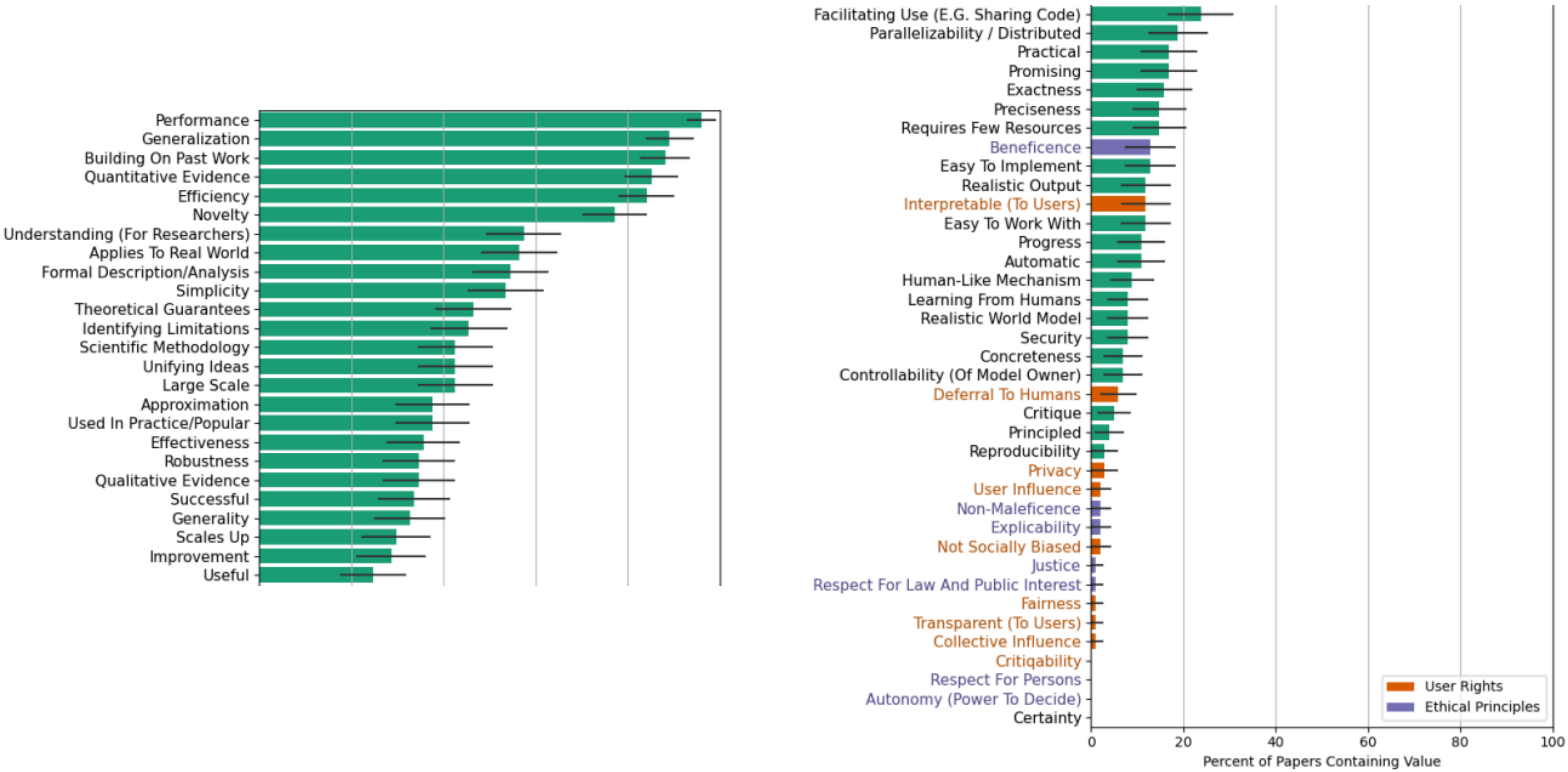
The truth is that any scientific discipline
When practiced properly
When studied rigorously
Is hard

Negative Potential

- Negative potential is rarely discussed, only 2 papers in the sample
- In those 2 papers, negative potential was mainly discussed hypothetically and generally, rather than discussing the specific negative potential of the presented model/concept
- Even in applications with well known negative potential such as surveillance and misinformation technologies do not discuss negative impacts



Discussion of Negative Potential	**% of Papers**
Does not mention negative potential	98%
Mentions but does not discuss negative potential	1%
Discusses negative potential	1%
Deepens our understanding of negative potential	0%


Performance
Generalization
Building On Past Work
Quantitative Evidence
Efficiency
Novelty
Understanding (For Researchers)
Applies To Real World
Formal Description/Analysis
Simplicity
Theoretical Guarantees
Identifying Limitations
Scientific Methodology
Unifying Ideas
Large Scale
Approximation
Used In Practice/Popular
Effectiveness
Robustness
Qualitative Evidence
Successful
Generality
Scales Up
Improvement
Useful
Facilitating Use (E.G. Sharing Code)
Parallelizability / Distributed
Practical
Promising
Exactness
Preciseness
Requires Few Resources
Beneficence
Easy To Implement
Realistic Output
Interpretable (To Users)
Easy To Work With
Progress
Automatic
Human-Like Mechanism
Learning From Humans
Realistic World Model
Security
Concreteness
Controllability (Of Model Owner)
Deferral To Humans
Critique
Principled
Reproducibility
Privacy
User Influence
Non-Maleficence
Explicability
Not Socially Biased
Justice
Respect For Law And Public Interest
Fairness
Transparent (To Users)
Collective Influence
Critiqability
Respect For Persons
Autonomy (Power To Decide)
Certainty
0
20
40
60
80
100
Percent of Papers Containing Value
User Rights
Ethical Principles

Top Values in ML Communities

- Performance (96% of papers),
- Generalization (89%),
- Building on past work (88%),
- Quantitative evidence (85%),
- Efficiency (84%)
- Novelty (77%)

Performance

- Typically measured as correctness averaged across predictions
    - Typically equal weighting for all predictions
    - Fairness research considers alternatives
- Pre-established large datasets with discrete ground truth labels are typically used for evaluation
    - Choice of dataset typically based on improving on baselines from prior work
    - Choice of dataset also justified by real world applicability
    - Using the same dataset encodes favoritism when the dataset isn’t accessible to all.
- Overall performance is highly valued
    - So much so that “success” and “improvement” become synonymous with accuracy

Generalization

- Shown by performing well on multiple tasks or multiple datasets
- The choice of datasets or choice of task indicating generalization are often not justified, unclear when performance will continue to generalize beyond carefully selected tasks or datasets
- Generalization generally is not with respect to more realistic applications
- Assuming generalizability often assumes new data should be treated similarly to previously seen data

Efficiency

- Efficiency commonly measured with respect to the ability to scale
- Not typically measured with respect to saving resources, in fact, papers that emphasize scalable ‘efficiency’ often require more resources than previous applications
- This incentivising scaling without consideration for resources incentives the most powerful actors to make increasingly inaccessible and reducing the ability to compete

Novelty & Building on Past Work

- Technical novelty is most valued
- As opposed to applying existing technology to a novel domain or novel philosophical argument
- Minimal critique of prior work
- Limitations of prior work are typically pointed out to emphasize the contributions of the current work
- Papers are rarely explicitly critical of prior work beyond pointing out these limitations
- Works that critique or rectify socially harmful aspects of prior work/datasets/goals are typically not seen as novel

- *In 2008/09, 24% of the top cited papers had corporate affiliated authors & in 2018/19 this statistic almost tripled to 71%.*
- *A concentration of a few large tech firms increasing more than fivefold, from 11% to 58%.*

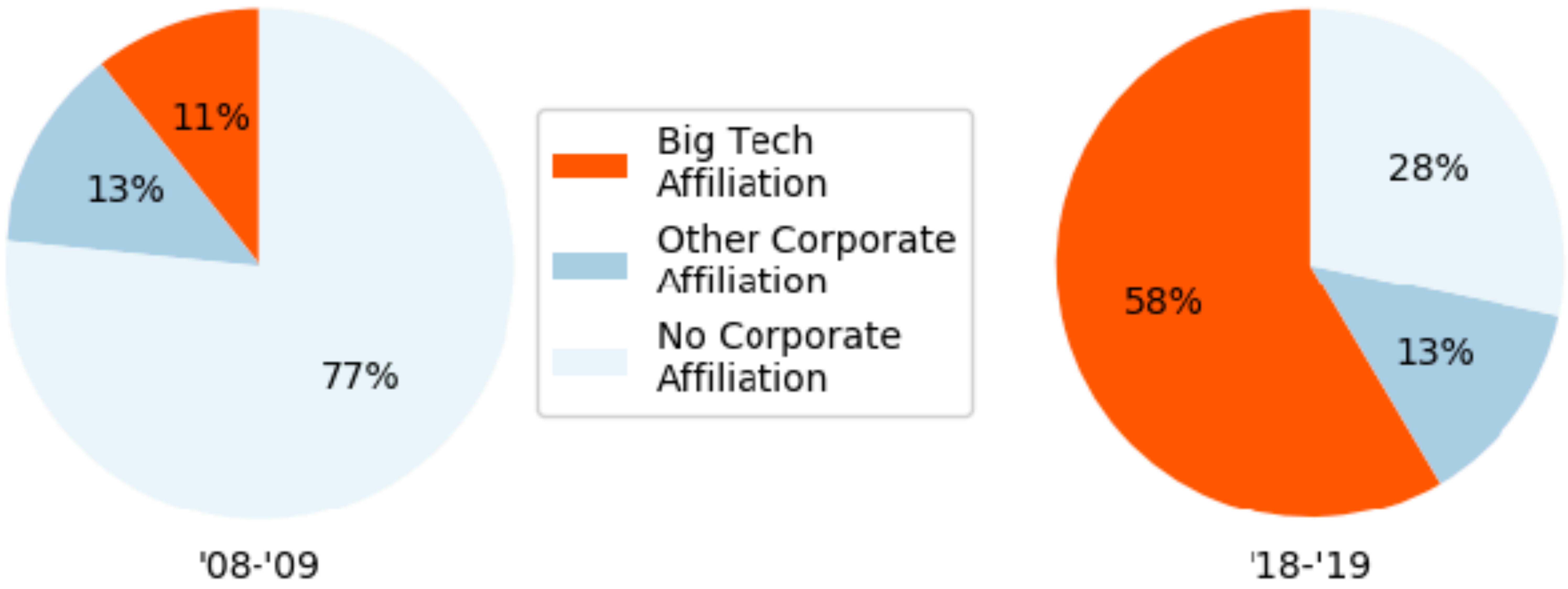


Figure 2: Corporate and Big Tech author affiliations.

User Rights and Ethical Values

Values like privacy, user input or influence, avoiding social biases, justice, and fairness are reflected in less than 10% of the papers analyzed

Negative potential of applications discussed in 1% (2 papers)

Positive Directions for User Rights and Ethical Values

Increase in value driven research organizations that center on ethical values

Increase in research funded by nonprofits

Increasing academic, nonprofit, and grassroots interest in upholding ethical values

Discussion Questions

1. Under what circumstances do the conventional values in the field contradict ethical values like privacy, justice, and combating social bias?
2. What are the positive and negative implications of the increasing influence of tech companies of funding and conducting academic research in the field?
3. What are the implications of ~50 universities producing most of the most citing work?

Fireside Chat with Pratyusha Kalluri