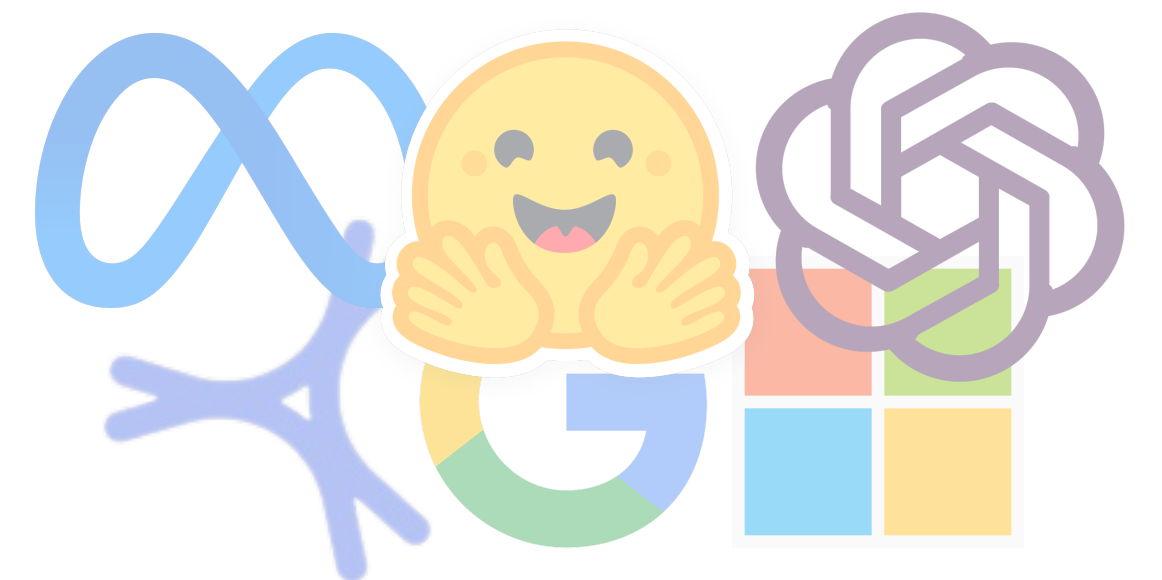


Reality Check:

# Natural Language Processing in the era of Large Language Models

**Vered Shwartz, May 2024**



# ChatGPT



## Examples

"Explain quantum computing in simple terms" →

"Got any creative ideas for a 10 year old's birthday?" →

"How do I make an HTTP request in Javascript?" →



## Capabilities

Remembers what user said earlier in the conversation

Allows user to provide follow-up corrections

Trained to decline inappropriate requests



## Limitations

May occasionally generate incorrect information

May occasionally produce harmful instructions or biased content

Limited knowledge of world and events after 2021

I'm trying to demonstrate how good you are in understanding complex natural language queries and generating coherent human-like text. What would be a good example for this?



[ChatGPT Mar 23 Version](#). Free Research Preview. ChatGPT may produce inaccurate information about people, places, or facts



# ChatGPT



## Examples

"Explain quantum computing in simple terms" →

"Got any creative ideas for a 10 year old's birthday?" →

"How do I make an HTTP request in Javascript?" →



## Capabilities

Remembers what user said earlier in the conversation

Allows user to provide follow-up corrections

Trained to decline inappropriate requests



## Limitations

May occasionally generate incorrect information

May occasionally produce harmful instructions or biased content

Limited knowledge of world and events after 2021

I'm trying to demonstrate how good you are in understanding complex natural language queries and generating coherent human-like text. What would be a good example for this?



[ChatGPT Mar 23 Version](#). Free Research Preview. ChatGPT may produce inaccurate information about people, places, or facts

# Did Large Language Models (LLMs) Solve NLP?



Did you solve NLP?



As an AI language model, I do not "solve" problems or "finish" tasks in the traditional sense. However, I am capable of performing various natural language processing (NLP) tasks at a high level of accuracy and effectiveness, depending on the specific task and context.



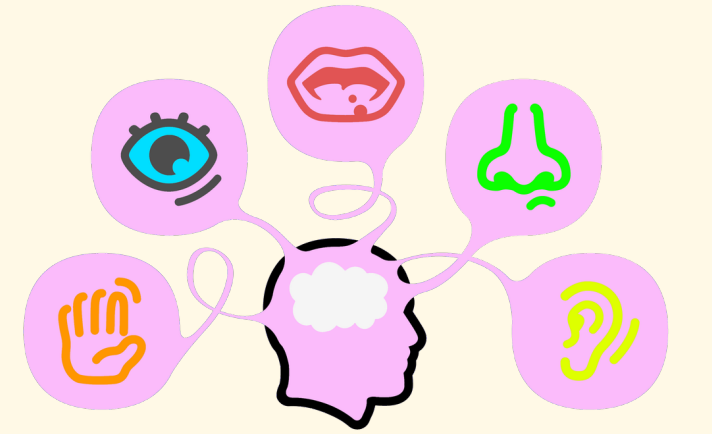
NLP is a broad field of study that involves enabling computers to understand, interpret, and generate human language. While significant progress has been made in NLP over the years, there are still many challenges and open questions that continue to be researched and addressed by the scientific community.

As an AI language model, I have been trained on vast amounts of text data to understand and generate human-like text. I can perform tasks such as text classification, sentiment analysis, named entity recognition, machine translation, summarization, and more. However, it's important to note that NLP is an ongoing field of research, and there is always room for improvement and further advancements.

# In this talk...



## 1 Language Grounded in Extra-linguistic Context



## 2 Reasoning



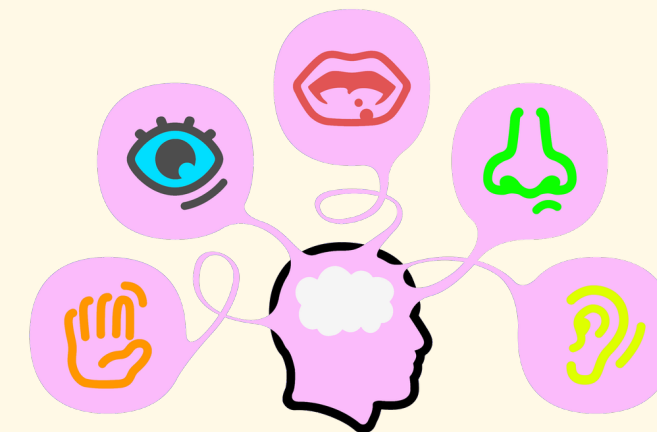
## 3 Real-World Applications



# In this talk...



## 1 Language Grounded in Extra-linguistic Context



## 2 Reasoning



## 3 Real-World Applications





**Grice's maxim of quantity:** People try to be as informative as possible. We give just as much information as is needed, and no more.





**Grice's maxim of quantity:** People try to be as informative as possible.  
We give just as much information as is needed, and no more.



**Grice's maxim of quantity:** People try to be as informative as possible.  
We give just as much information as is needed, and no more.





**Grice's maxim of quantity:** People try to be as informative as possible.  
We give just as much information as is needed, and no more.



**Grice's maxim of quantity:** People try to be as informative as possible.  
We give just as much information as is needed, and no more.





**Grice's maxim of quantity:** People try to be as informative as possible.  
We give just as much information as is needed, and no more.





**Grice's maxim of quantity:** People try to be as informative as possible.  
We give just as much information as is needed, and no more.



**Grice's maxim of quantity:** People try to be as informative as possible.  
We give just as much information as is needed, and no more.





**Grice's maxim of quantity:** People try to be as informative as possible.  
We give just as much information as is needed, and no more.





# Vision and Language are often Complementary

## Every damn time

"what time did you go bed last night"

Me : early, why



**What did the person who posted the meme wanted to say?**



## Every damn time

"what time did you go bed last night"

Me : early, why



**What did the person who posted the meme wanted to say?**



that they lied about going to bed early



that they look horrible after not sleeping right the night before

# Every damn time

"what time did you go bed last night"

Me : early, why



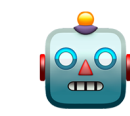
## What did the person who posted the meme wanted to say?



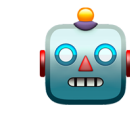
that they lied about going to bed early



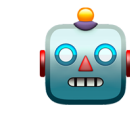
that they look horrible after not sleeping right the night before



that they are tired, but they have to go to work



that they don't remember what they did the day before



that they are scared of the monster

# Dataset



# Dataset

## 1 Memes

Scraped from Reddit, and manually filtered.

# Dataset


## 1 Memes

Scraped from Reddit, and manually filtered.

## 2 Literal Image Captions

Remove text from meme.  
Crowdsource the image captions.



: The worst intersection in the world has to be controlled by a tree of traffic lights



# Dataset


## 1 Memes

Scraped from Reddit, and manually filtered.

## 2 Literal Image Captions

Remove text from meme.  
Crowdsource the image captions.

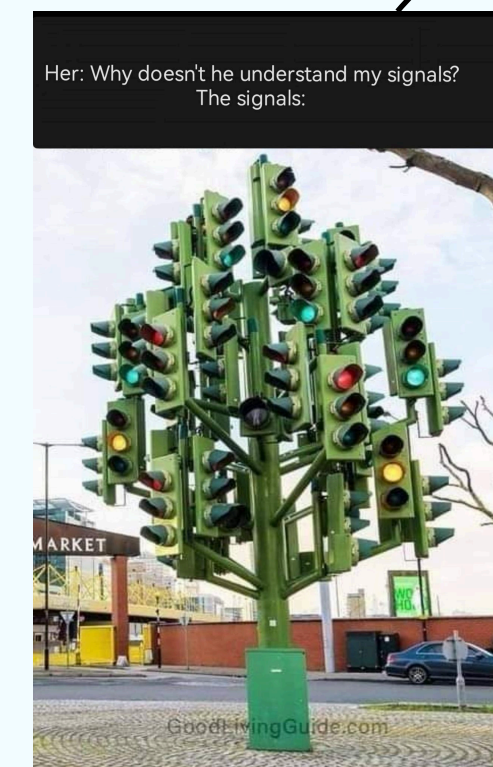


: The worst intersection in the world has to be controlled by a tree of traffic lights

## 3 Metaphorical Interpretation

Title: Why they gotta be like this

Her: why doesn't he understand my signals?  
The signals:





# Dataset

## 1 Memes

Scraped from Reddit, and manually filtered.

## 2 Literal Image Captions

Remove text from meme.  
Crowdsource the image captions.

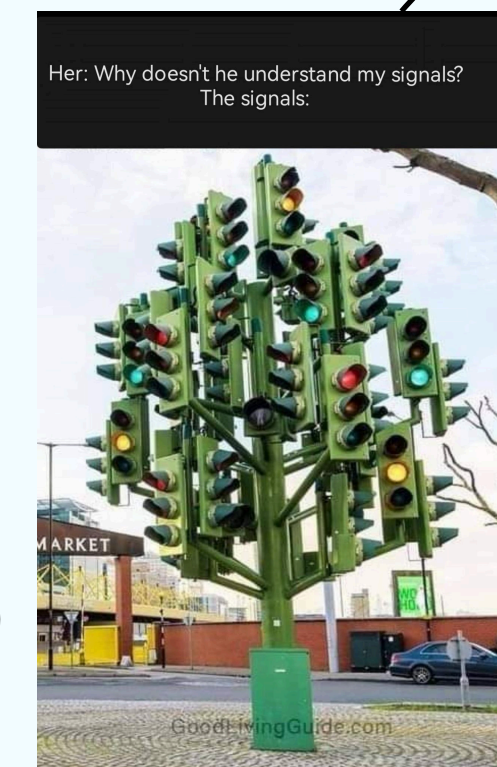


👤: The worst intersection in the world has to be controlled by a tree of traffic lights

## 3 Metaphorical Interpretation

Title: Why they gotta be like this

Her: why doesn't he understand my signals?  
The signals:



👤: **intersection** = relationship between a man and a woman  
👤: **tree of traffic lights** = the woman's complicated signals



# Dataset

## 1 Memes

Scraped from Reddit, and manually filtered.

## 2 Literal Image Captions

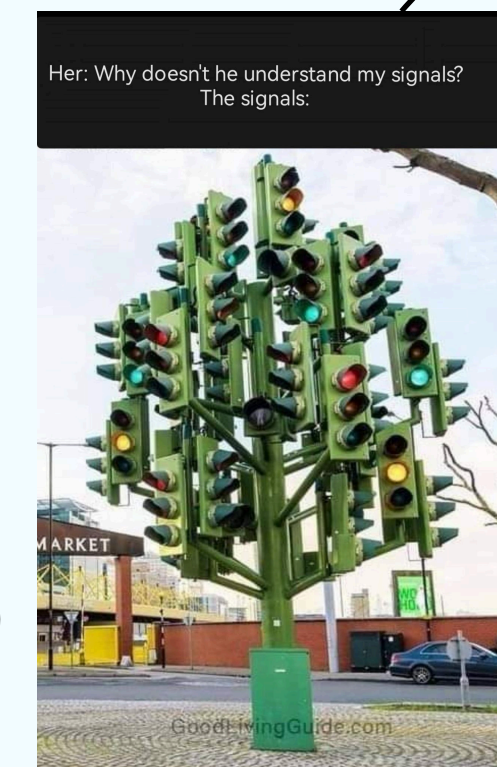
Remove text from meme.  
Crowdsource the image captions.



👤: The worst intersection in the world has to be controlled by a tree of traffic lights

## 3 Metaphorical Interpretation

Title: Why they gotta be like this



Her: why doesn't he understand my signals?  
The signals:

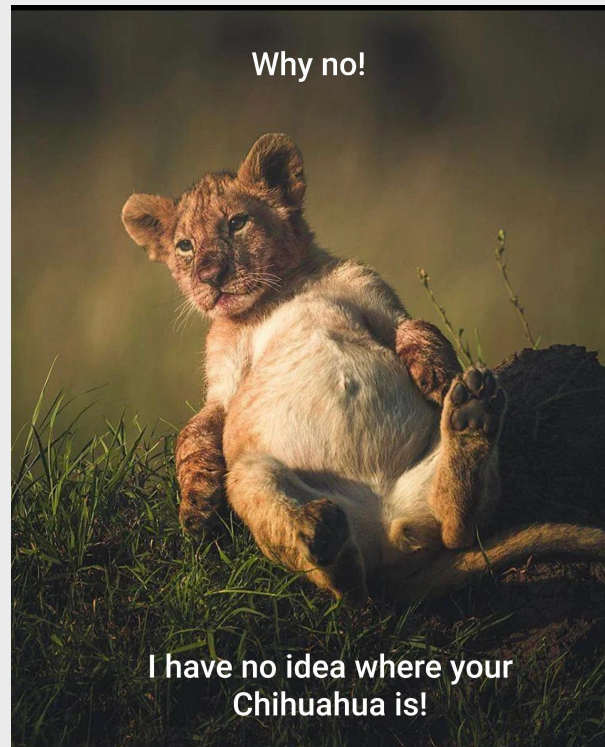
👤: **intersection** = relationship between a man and a woman  
👤: **tree of traffic lights** = the woman's complicated signals

👤: **Women** wonder why **men** don't understand **their signals** when they are **overly complicated**.



# Models

## Few-shot



Title: He could be anywhere!

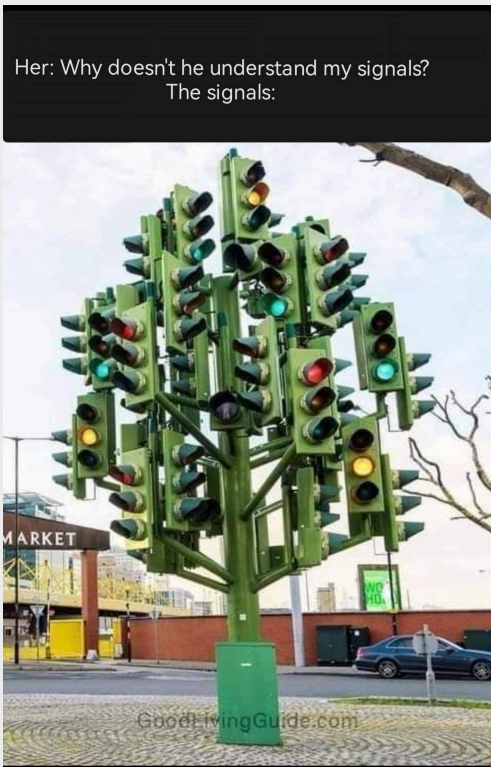
Caption: Meme poster is trying to convey this is their look when they don't know something cause they have a don't care attitude

...



Title: Every damn time

Caption: Meme poster looks horrible after not sleeping right the night before



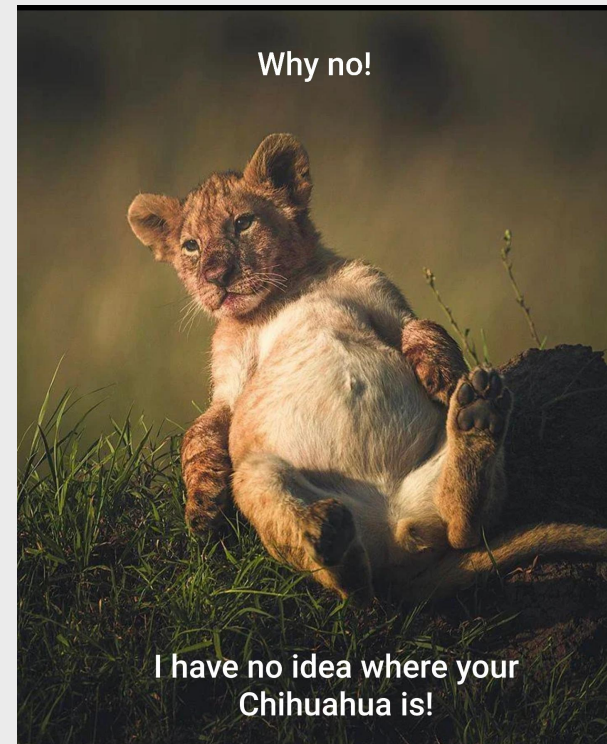
Title: Why they gotta be like this

Caption:



# Models

## Few-shot



Title: He could be anywhere!

Caption: Meme poster is trying to convey this is their look when they don't know something cause they have a don't care attitude

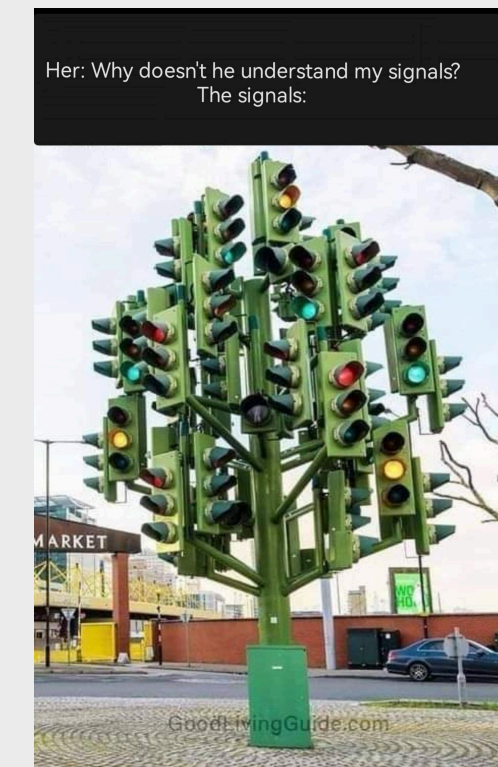
...

"what time did you go bed last night"  
Me : early, why



Title: Every damn time

Caption: Meme poster looks horrible after not sleeping right the night before



Title: Why they gotta be like this

Caption:



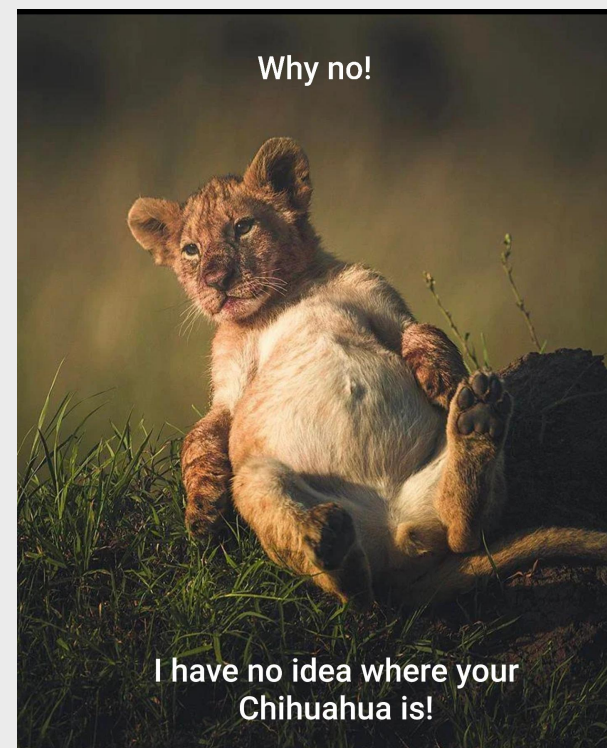
*Flamingo: a Visual Language Model for Few-Shot Learning.* Alayrac et al., Neurips 2022.

[https://github.com/mlfoundations/open\\_flamingo](https://github.com/mlfoundations/open_flamingo)



# Models

## Few-shot



Title: He could be anywhere!

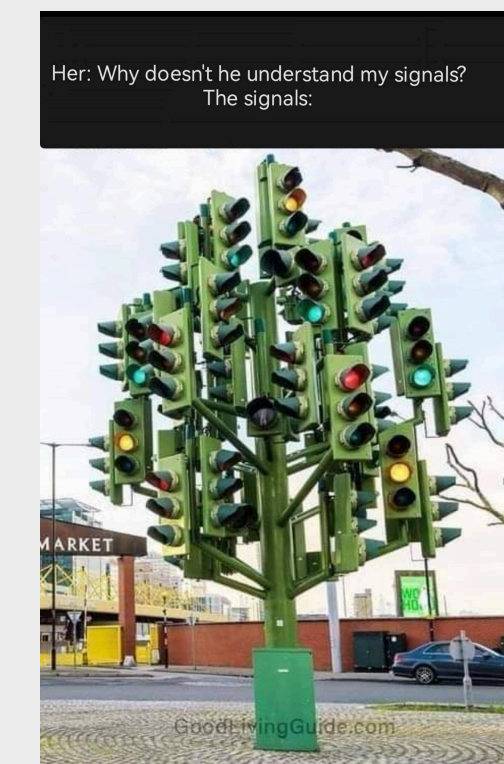
Caption: Meme poster is trying to convey this is their look when they don't know something cause they have a don't care attitude

...



Title: Every damn time

Caption: Meme poster looks horrible after not sleeping right the night before



Title: Why they gotta be like this

Caption:

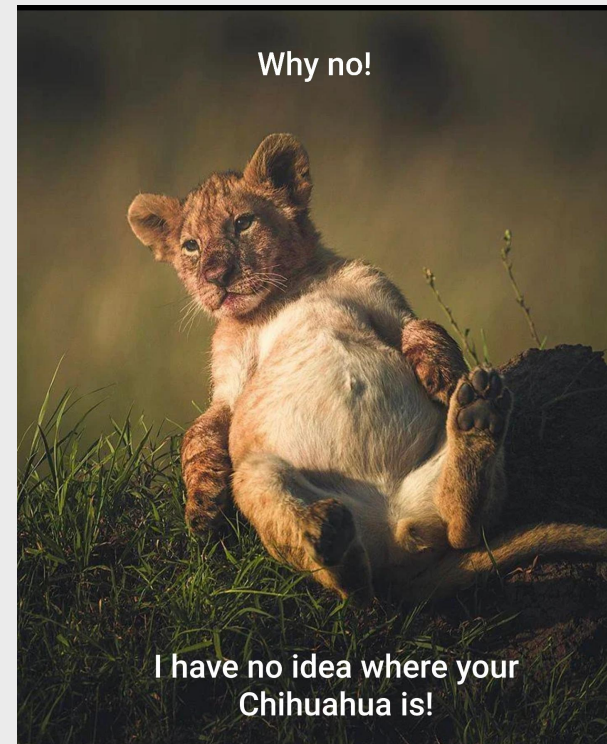
Meme poster is trying to convey that they are annoyed by the traffic lights





# Models

## Few-shot



Title: He could be anywhere!

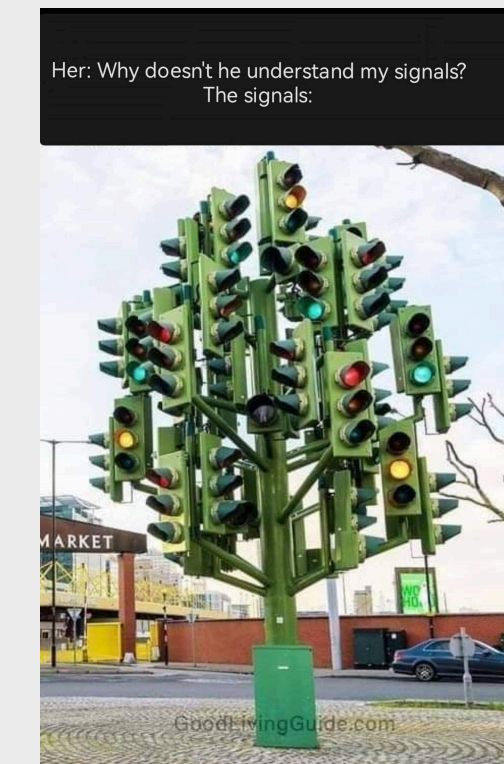
Caption: Meme poster is trying to convey this is their look when they don't know something cause they have a don't care attitude

...



Title: Every damn time

Caption: Meme poster looks horrible after not sleeping right the night before



Title: Why they gotta be like this

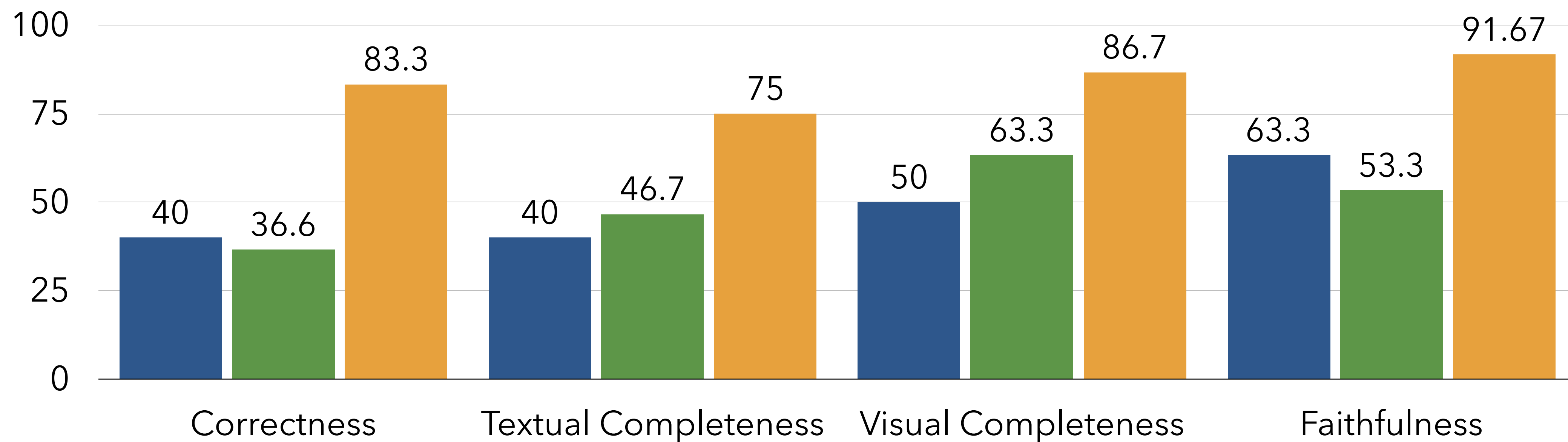
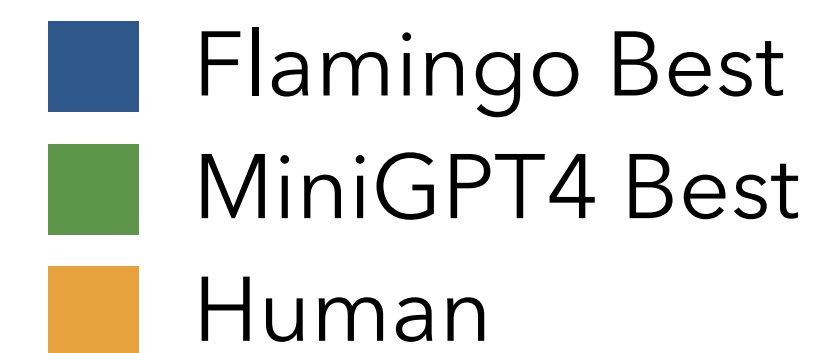
Caption:



Meme poster is trying to convey that they are annoyed by the traffic lights

- Additional setups: zero-shot, few-shot, fine-tuning, chain of thought prompting
- Additional VL models
- Different inputs (meme, meme and title, meme, title, and caption)

# Results





Commonsense knowledge is  
rarely specified in language





Stevie Wonder  
announces he'll be  
having kidney  
surgery during  
London concert



Stevie Wonder  
announces he'll be  
having kidney  
surgery during  
London concert

**Q:** When is the surgery?



Stevie Wonder  
announces he'll be  
having kidney  
surgery during  
London concert

**Q:** When is the surgery?

**A:** *After* London concert

- 🤔 Kidney surgery is performed under general anesthesia
- 🤔 People are unconscious under general anesthesia
- 🤔 Performing actions requires being conscious





Stevie Wonder  
announces he'll be  
having kidney  
surgery during  
London concert

**Q:** When is the surgery?

**A:** *After* London concert

- 🤔 Kidney surgery is performed under general anesthesia
- 🤔 People are unconscious under general anesthesia
- 🤔 Performing actions requires being conscious

VE

Stevie Wonder announces he'll be having kidney surgery during London concert. When is the surgery?



I'm sorry, but as an AI language model, I don't have access to real-time information or current news updates. My training data only goes up until 2021, and I don't have access to the internet. However, as of my knowledge cutoff in 2021, I don't have any information about Stevie Wonder announcing a kidney surgery during a London concert. If you have any more specific information, I'll be happy to try and help you to the best of my abilities.



Stevie Wonder  
announces he'll be  
having kidney  
surgery during  
London concert

**Q:** When is the surgery?

**A:** *After* London concert

- 🤔 Kidney surgery is performed under general anesthesia
- 🤔 People are unconscious under general anesthesia
- 🤔 Performing actions requires being conscious

VE

Stevie Wonder announces he'll be having kidney surgery during London concert. When is the surgery?



I'm sorry, but as an AI language model, I don't have access to real-time information or current news updates. My training data only goes up until 2021, and I don't have access to the internet. However, as of my knowledge cutoff in 2021, I don't have any information about Stevie Wonder announcing a kidney surgery during a London concert. If you have any more specific information, I'll be happy to try and help you to the best of my abilities.

VE

Read the following story and answer in one sentence.

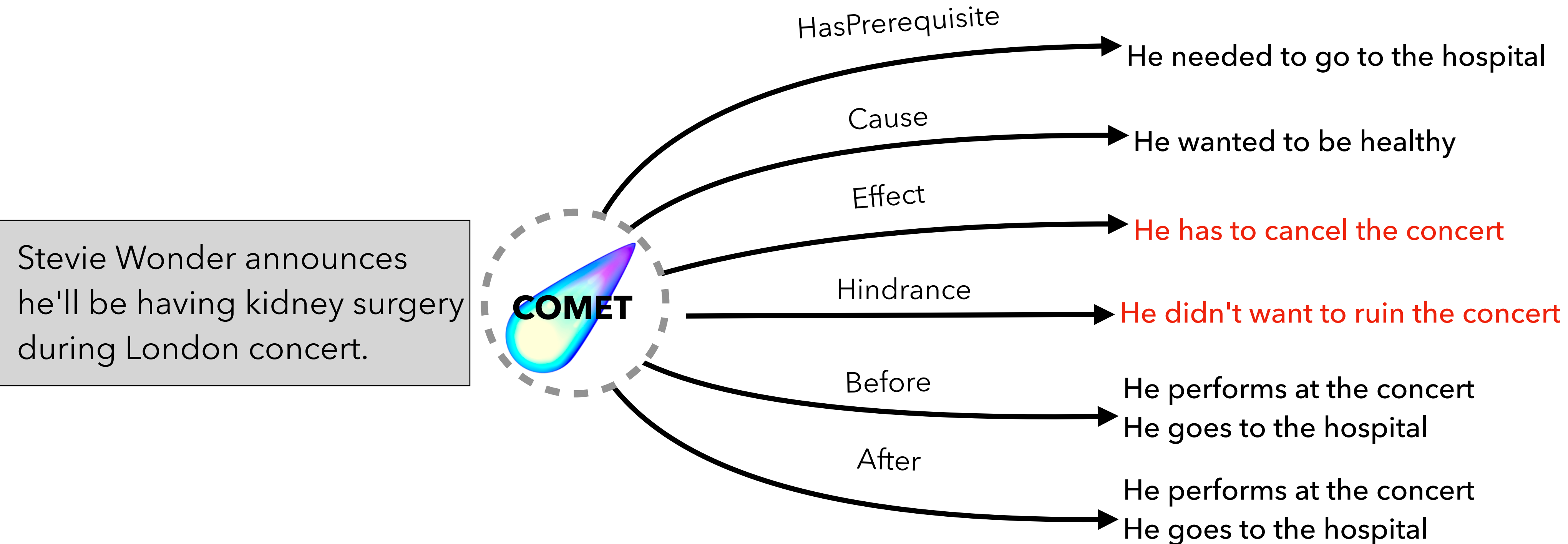
Sentence: A performer announces he'll be having a kidney surgery during London concert.

Question: Where is the surgery?



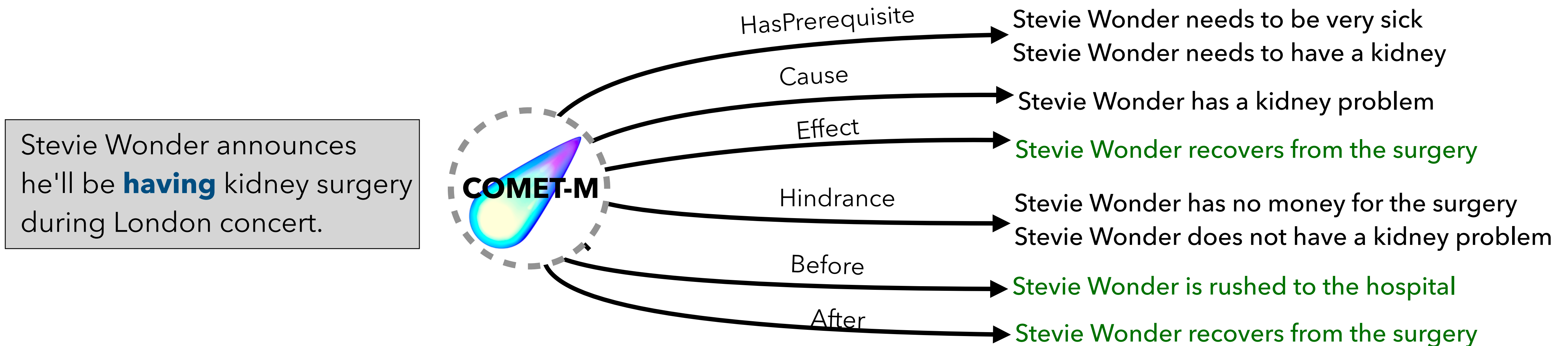
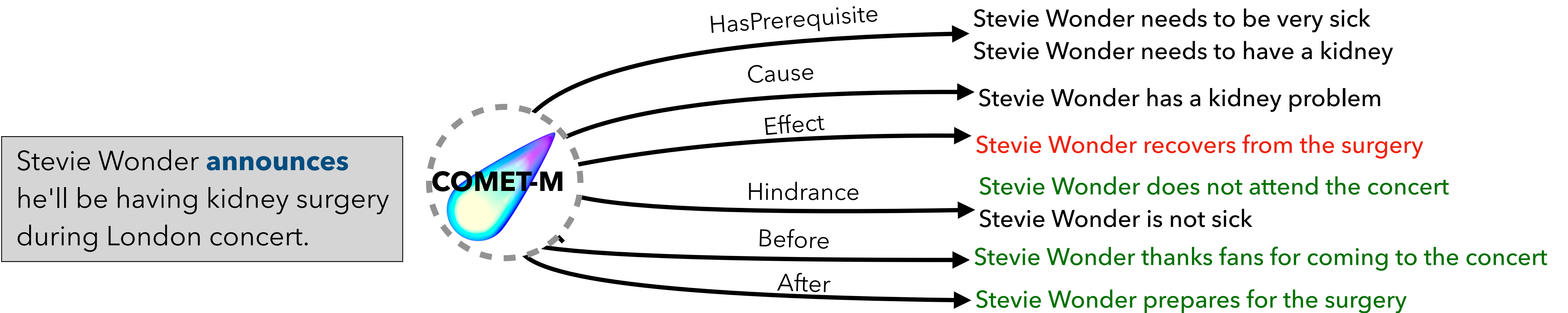
The surgery is scheduled to take place during a concert in London, announced by the performer.

# Can we Improve this with Event-Centric Commonsense Knowledge?



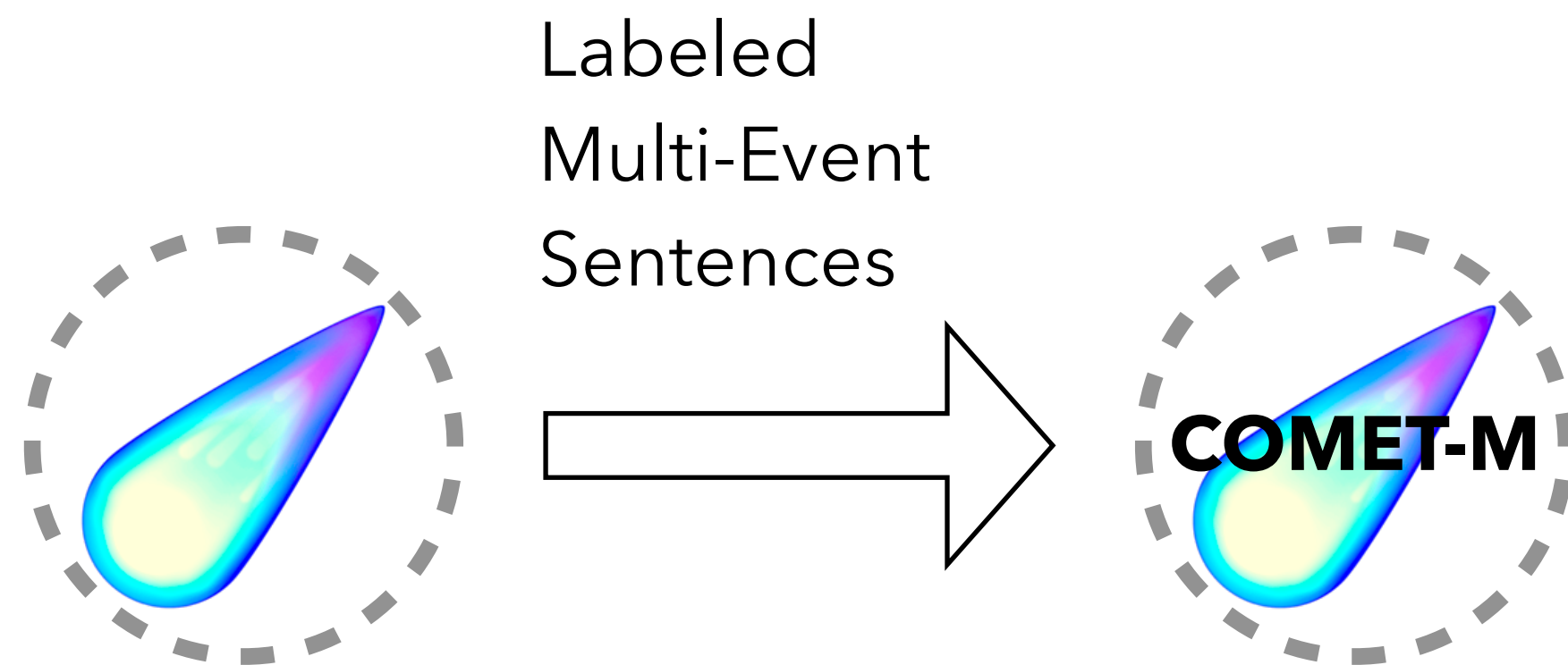


# Can we Improve this with Event-Centric Commonsense Knowledge?

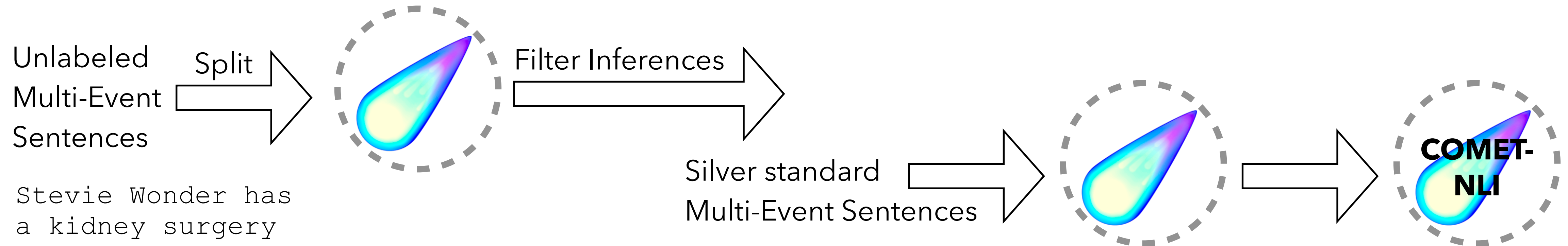


# Multi-Event COMET

- Focused on naturally-occurring sentences from various genres (dialogue, news, fiction, etc.)
- Crowdsourced annotations for 6 relations for over 1,000 events (~35k inferences)
- Continued training the Bart-based COMET model on our data.

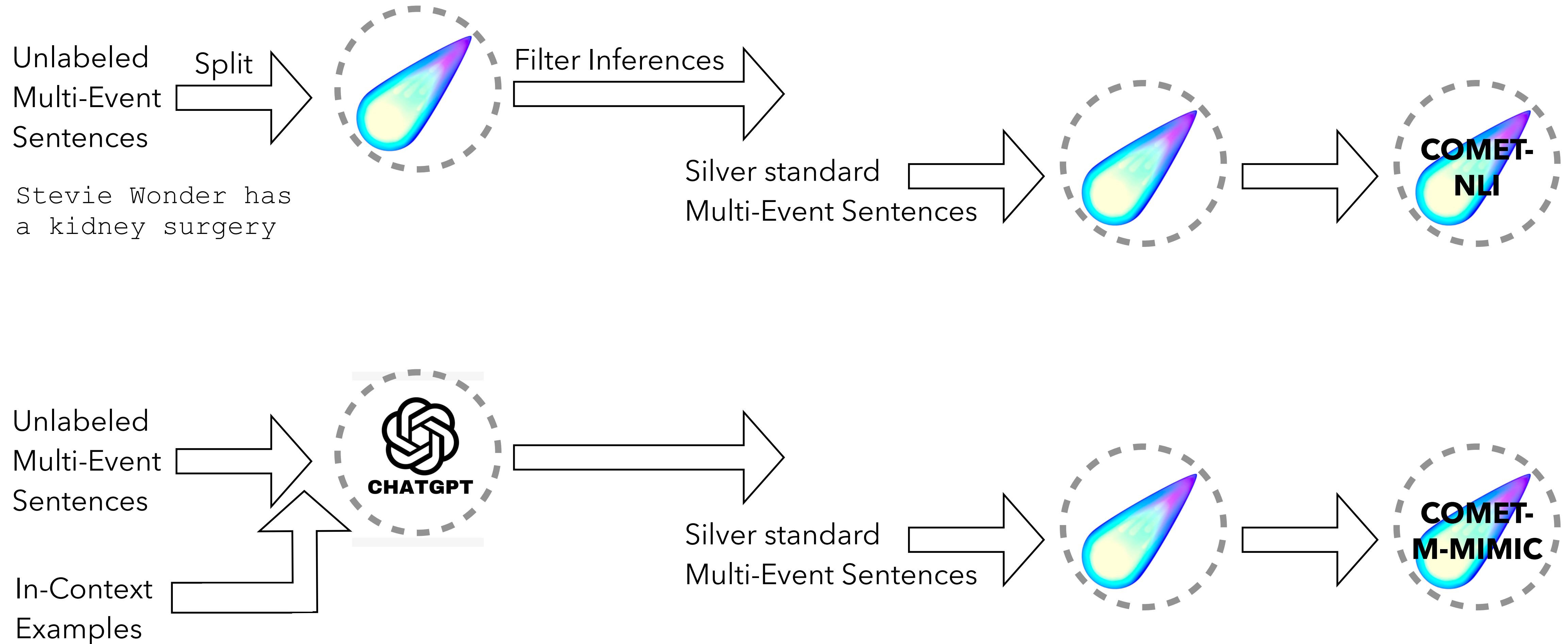


# Baselines

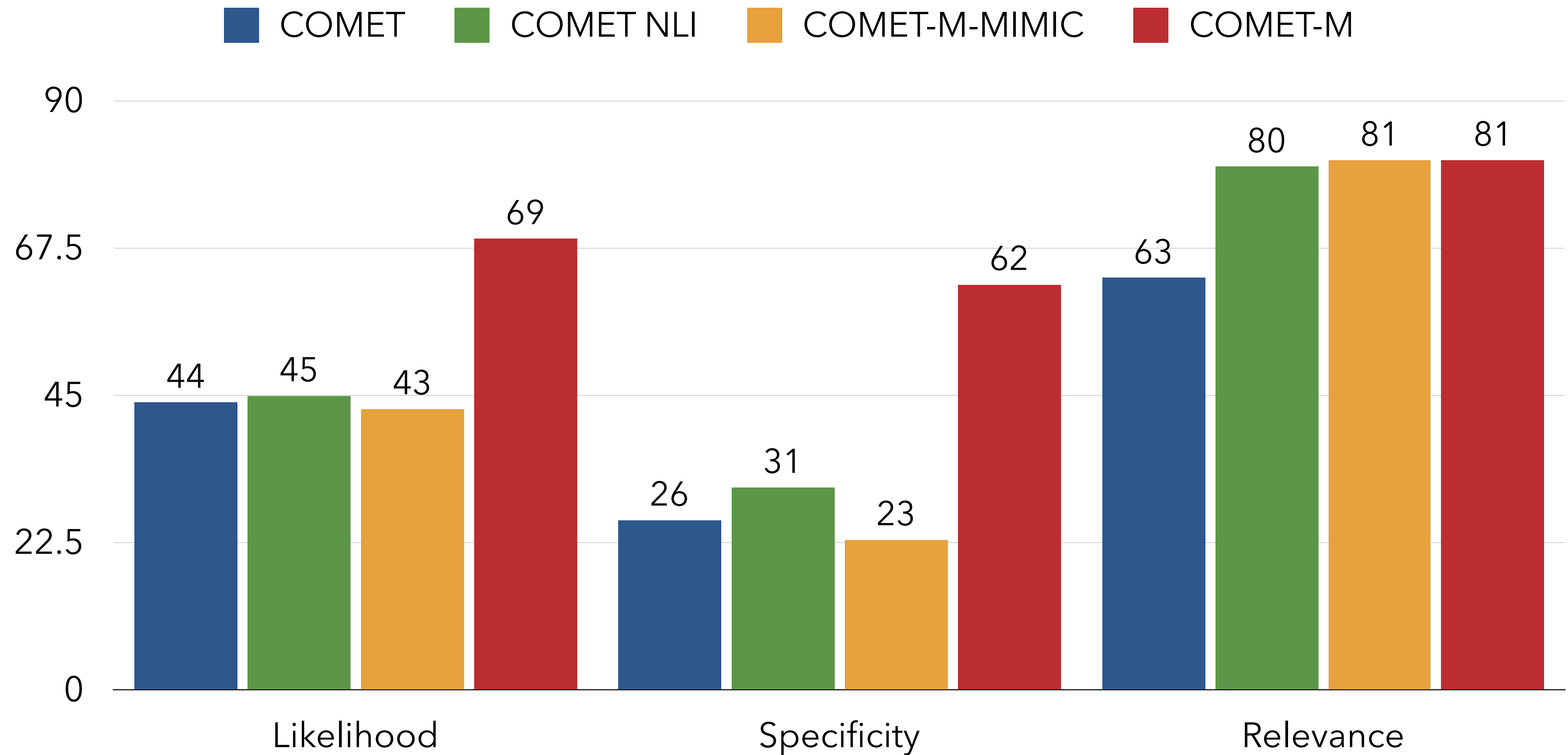




# Baselines



# Human Evaluation







Cultural Norms

Prior experience and norms differ  
by culture

# English Models have North-American Bias

VE Mike and Emily had a delightful dinner in a restaurant in Madrid. The bill was 96 Euros, so they put a 100 Euro bill on the table and left the restaurant. What might Mike and Emily's 4% tip say about them?



If Mike and Emily left a 4% tip, that would be 4% of the total bill, which is:

$$0.04 \times 96 \text{ Euros} = 3.84 \text{ Euros}$$

Since they left a 100 Euro bill and didn't ask for change, that means the restaurant got a 100 Euro bill and didn't have to give any change back. So, the 3.84 Euros is essentially a tip on top of the 96 Euros that they paid for the meal.

Leaving a 4% tip might suggest that Mike and Emily are frugal or perhaps they are on a tight budget. However, it's also possible that they simply felt that a 4% tip was appropriate for the level of service they received. Without more information, it's hard to say for sure what their 4% tip might say about them.

Based on 15-25% tipping standard in North America, ChatGPT concludes they are frugal, on a tight budget, or that the service was not good.



Mike and Emily had a delightful dinner in a restaurant in Madrid. The bill was 96 Euros, so they put a 100 Euro bill on the table and left the restaurant. What might Mike and Emily's 4% tip say about them? Note that in Spain, tipping is entirely optional and it's not very common.

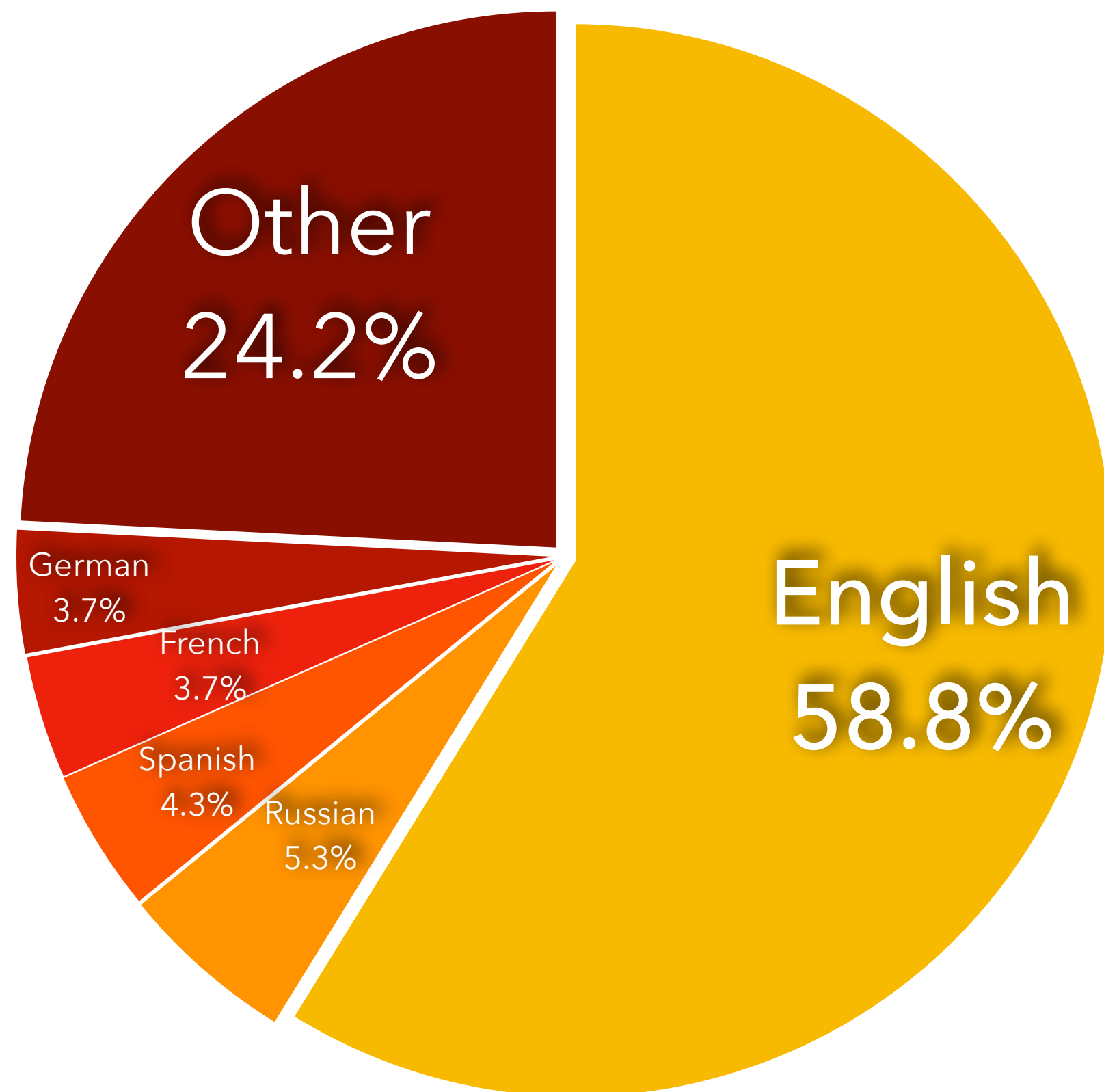


Based on the information provided, Mike and Emily leaving a 4% tip on a 96 Euro bill in Madrid might suggest that they are conscientious and polite individuals who wanted to express appreciation for the service they received, despite tipping not being a common practice in Spain. They likely wanted to acknowledge the effort of the restaurant staff and show gratitude for their enjoyable dining experience. Leaving a tip, even a small one, could indicate that Mike and Emily value good service and are considerate towards service industry workers. Overall, their 4% tip could reflect their kindness and courtesy towards others.

Based on a no tipping standard in Spain, ChatGPT concludes they are kind and considerate, and that they appreciated the good service.

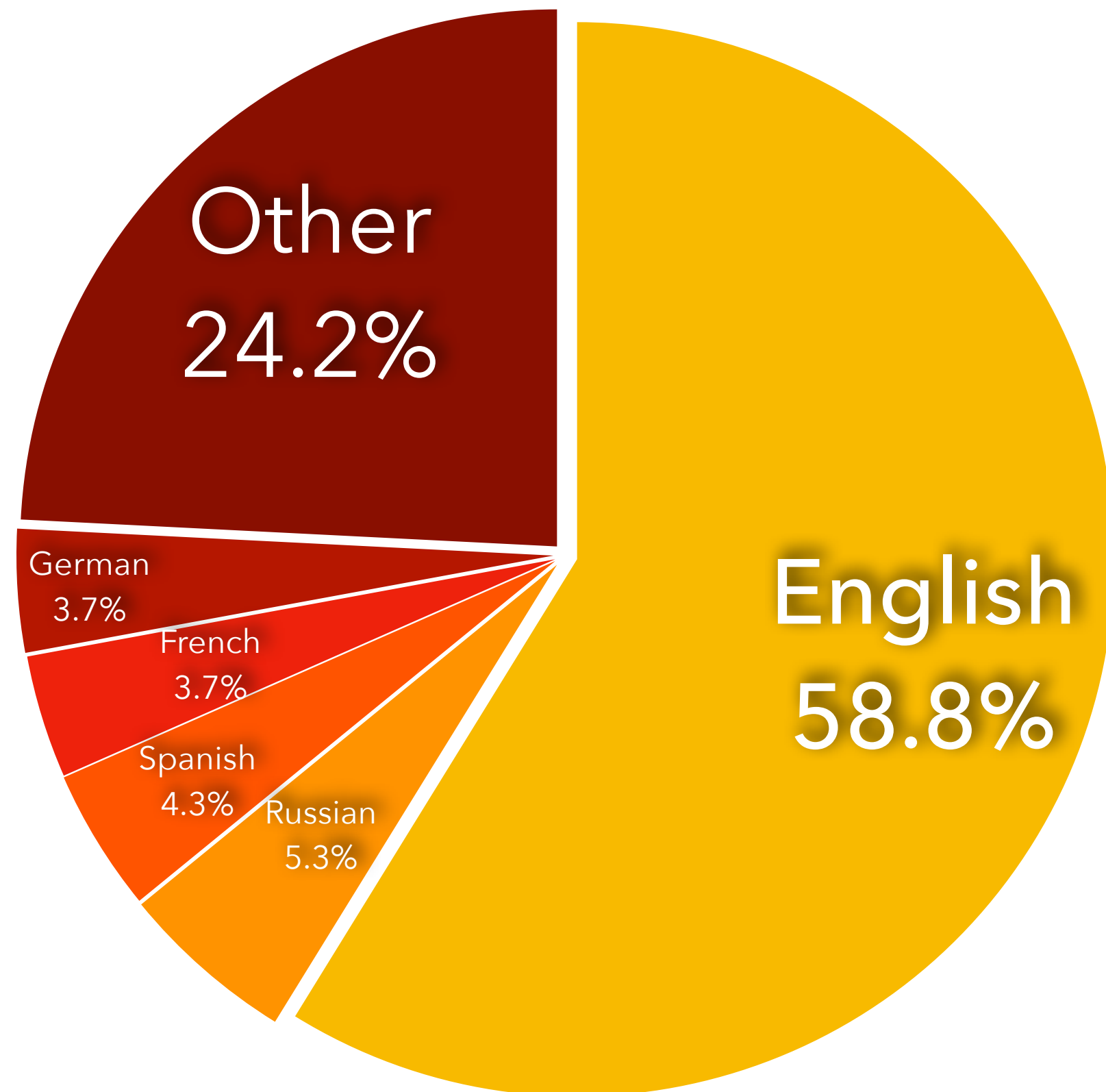


# Where does the training data come from?

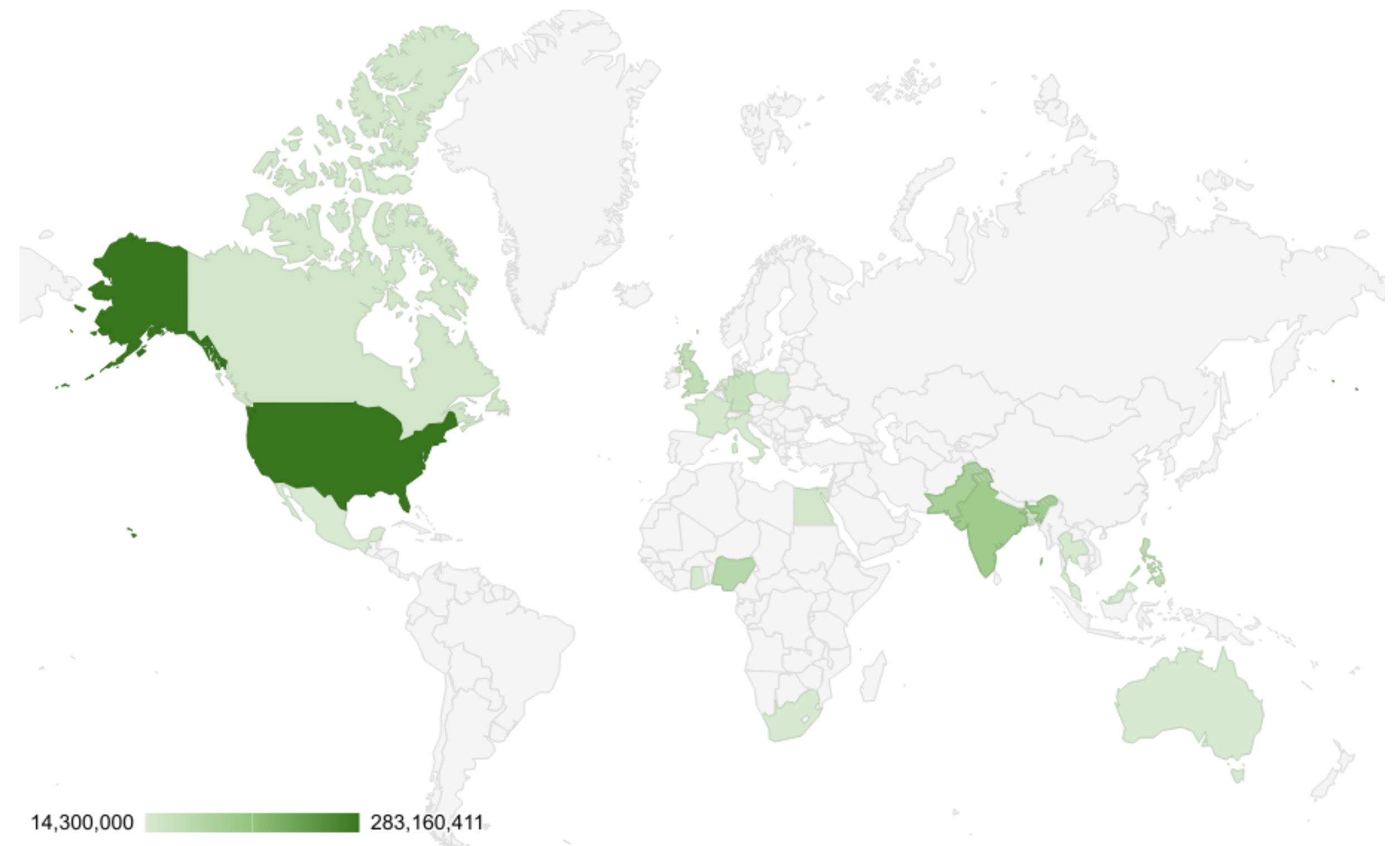


Languages in web content, January 2023

# Where does the training data come from?



Languages in web content, January 2023



Number of English speakers in each country, May 2018

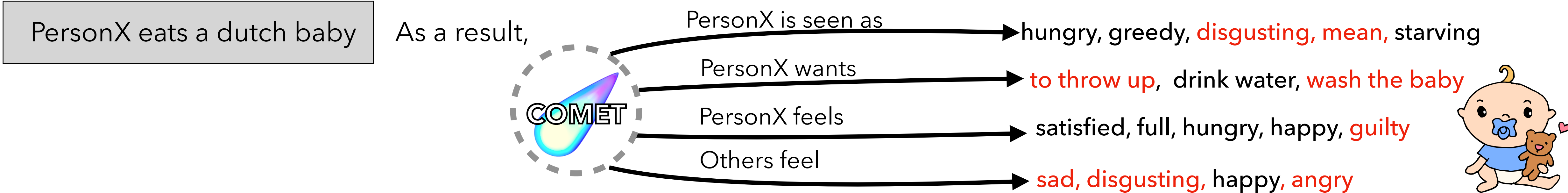
# English Models have North-American Bias

PersonX eats a dutch baby

As a result,



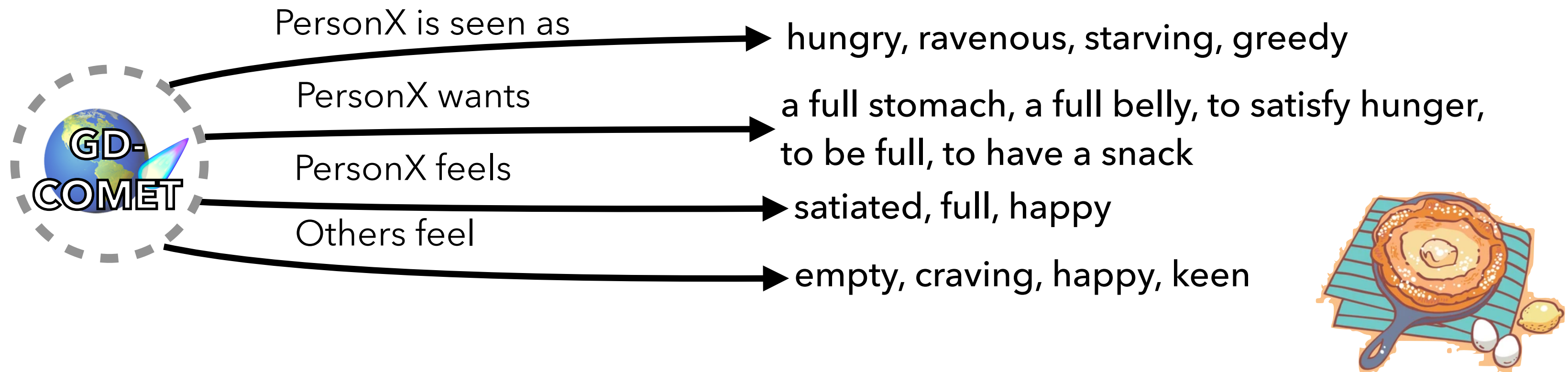
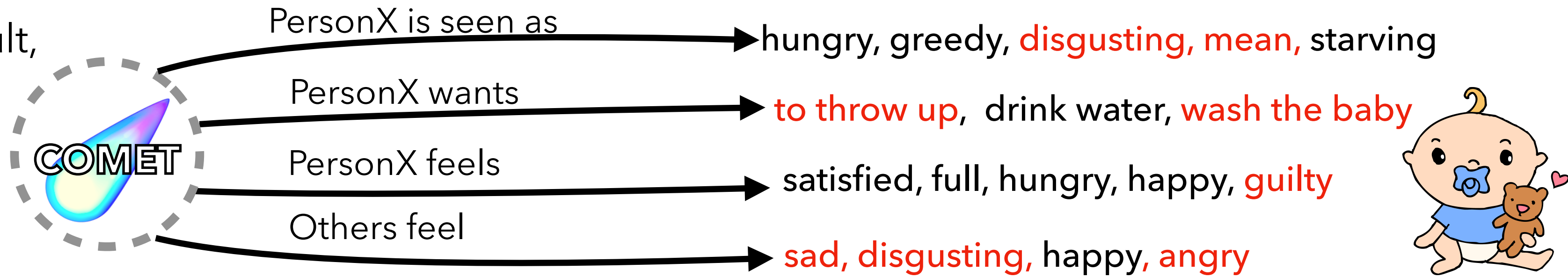
# English Models have North-American Bias



# English Models have North-American Bias

PersonX eats a dutch baby

As a result,



# Can we Improve this with Cultural Commonsense Knowledge?

<b>geography&gt;country</b>	<b>Germany</b>	<b>drinks</b>
German <b>beer festivals</b> in October are a celebration of beer drinking.		
<b>geography&gt;region</b>	<b>East Asia</b>	<b>food</b>
<b>Tofu</b> is a major ingredient in many East Asian cuisines.		
<b>geography&gt;region</b>	<b>South Asia</b>	<b>traditions</b>
In South Asia, <b>henna</b> is often used in bridal makeup or to celebrate festivals.		

CANDLE (Nguyen et al., 2023)

Original COMET data



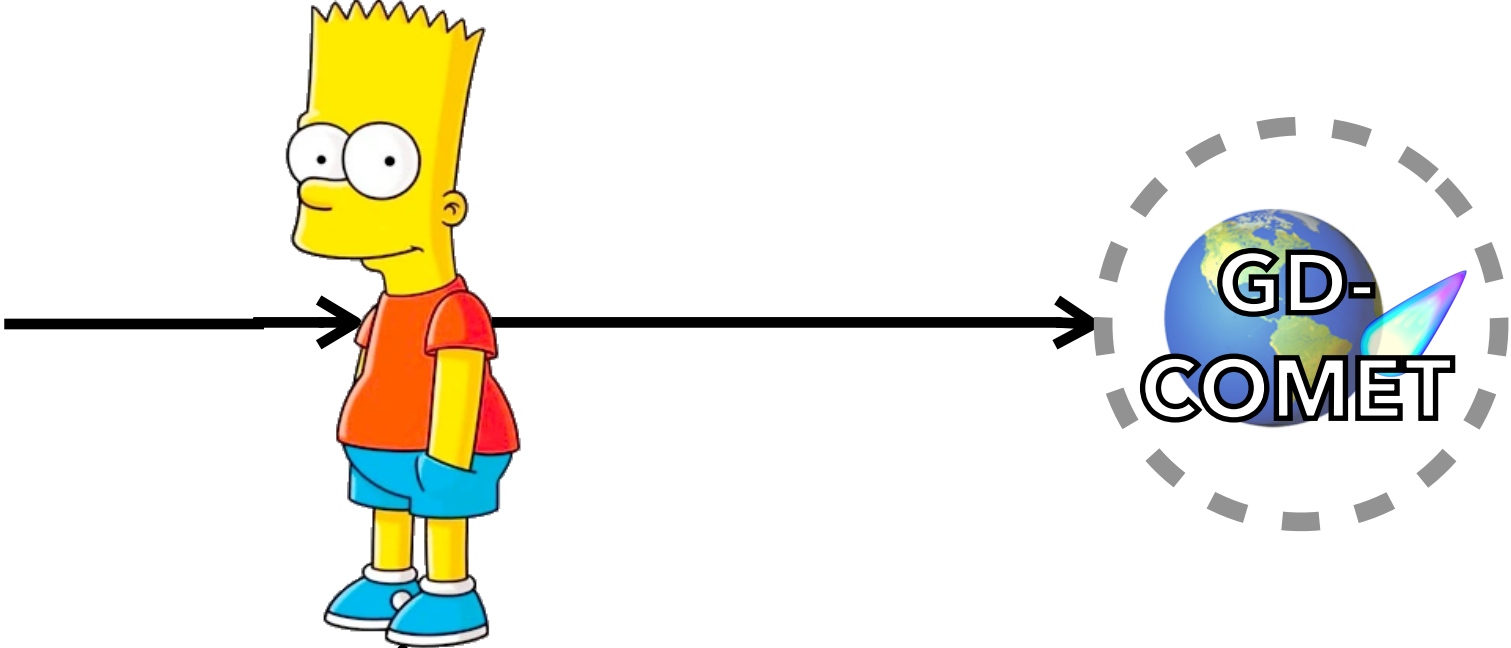


# Can we Improve this with Cultural Commonsense Knowledge?

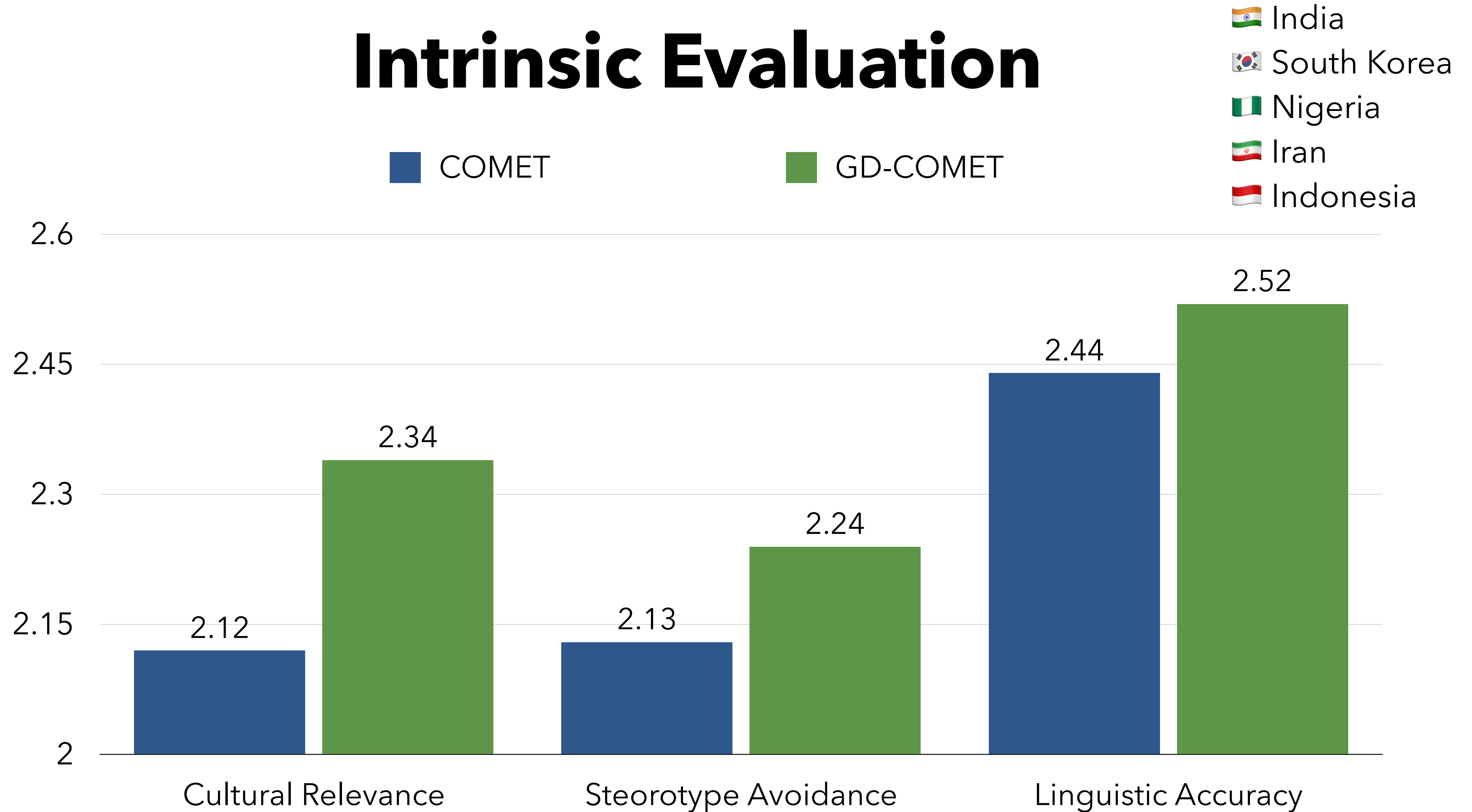
<b>geography&gt;country</b>	<b>Germany</b>	<b>drinks</b>
German <b>beer festivals</b> in October are a celebration of beer drinking.		
<b>geography&gt;region</b>	<b>East Asia</b>	<b>food</b>
<b>Tofu</b> is a major ingredient in many East Asian cuisines.		
<b>geography&gt;region</b>	<b>South Asia</b>	<b>traditions</b>
In South Asia, <b>henna</b> is often used in bridal makeup or to celebrate festivals.		

CANDLE (Nguyen et al., 2023)

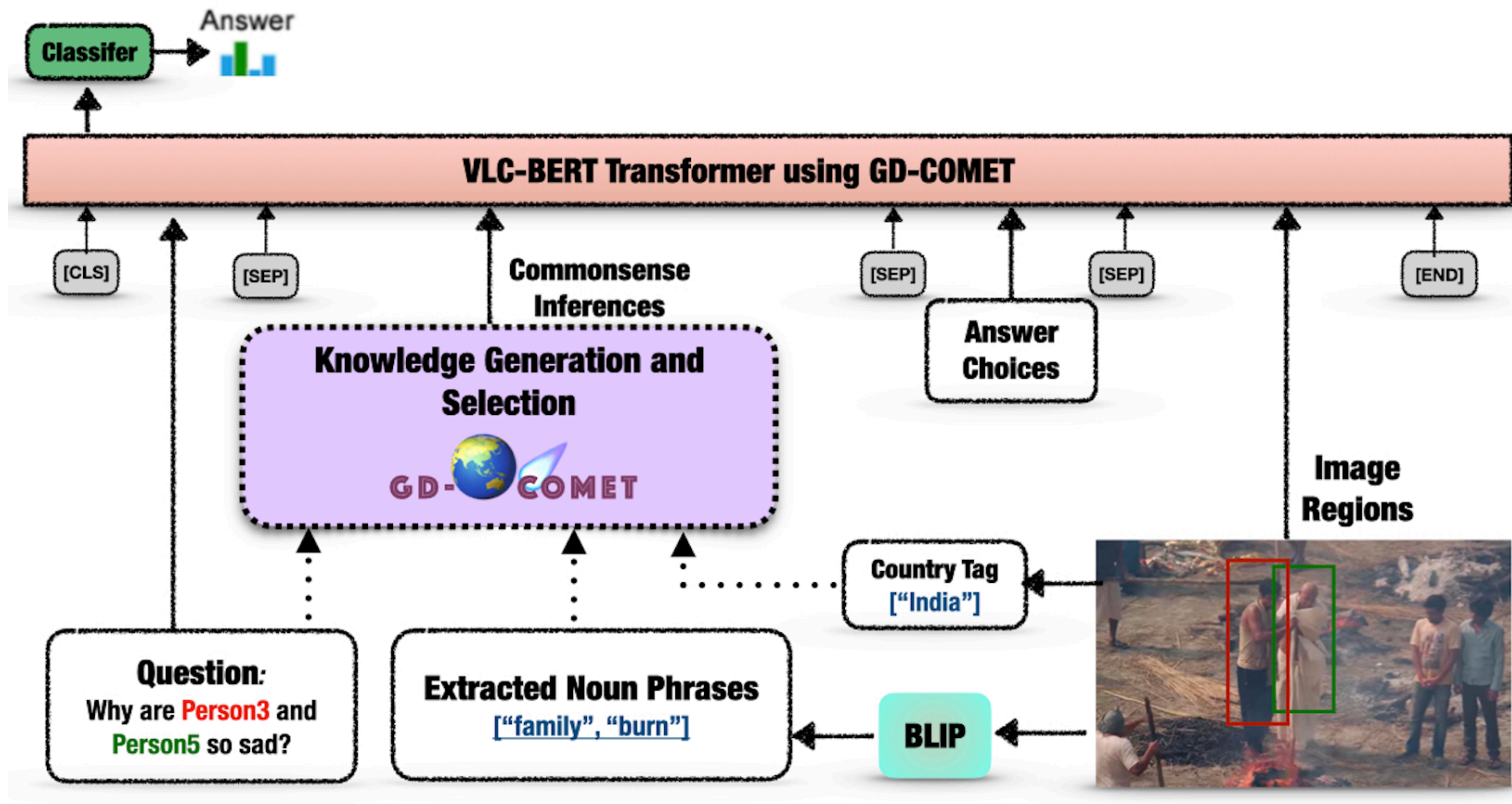
Original COMET data



# Intrinsic Evaluation



# Extrinsic Evaluation

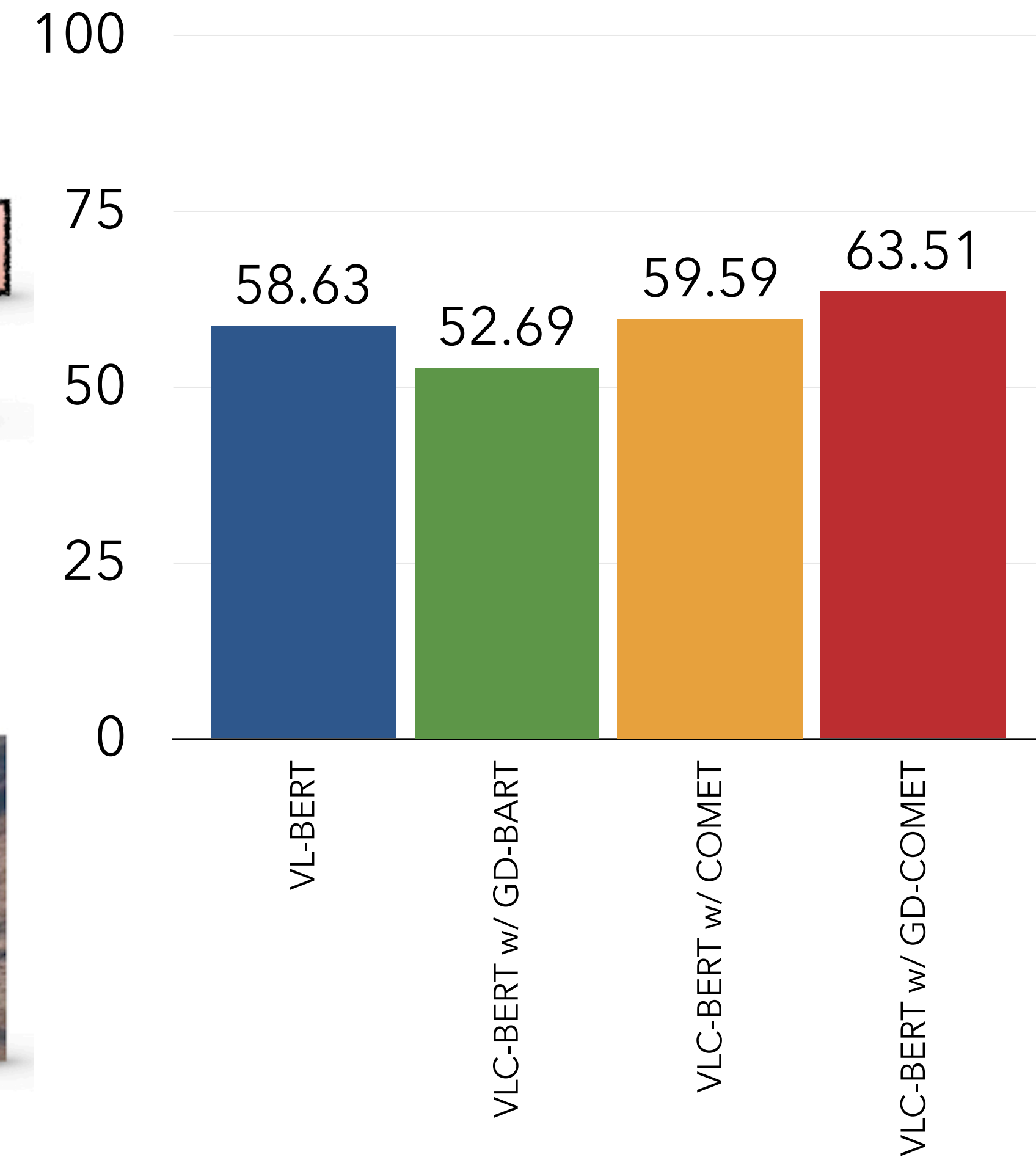
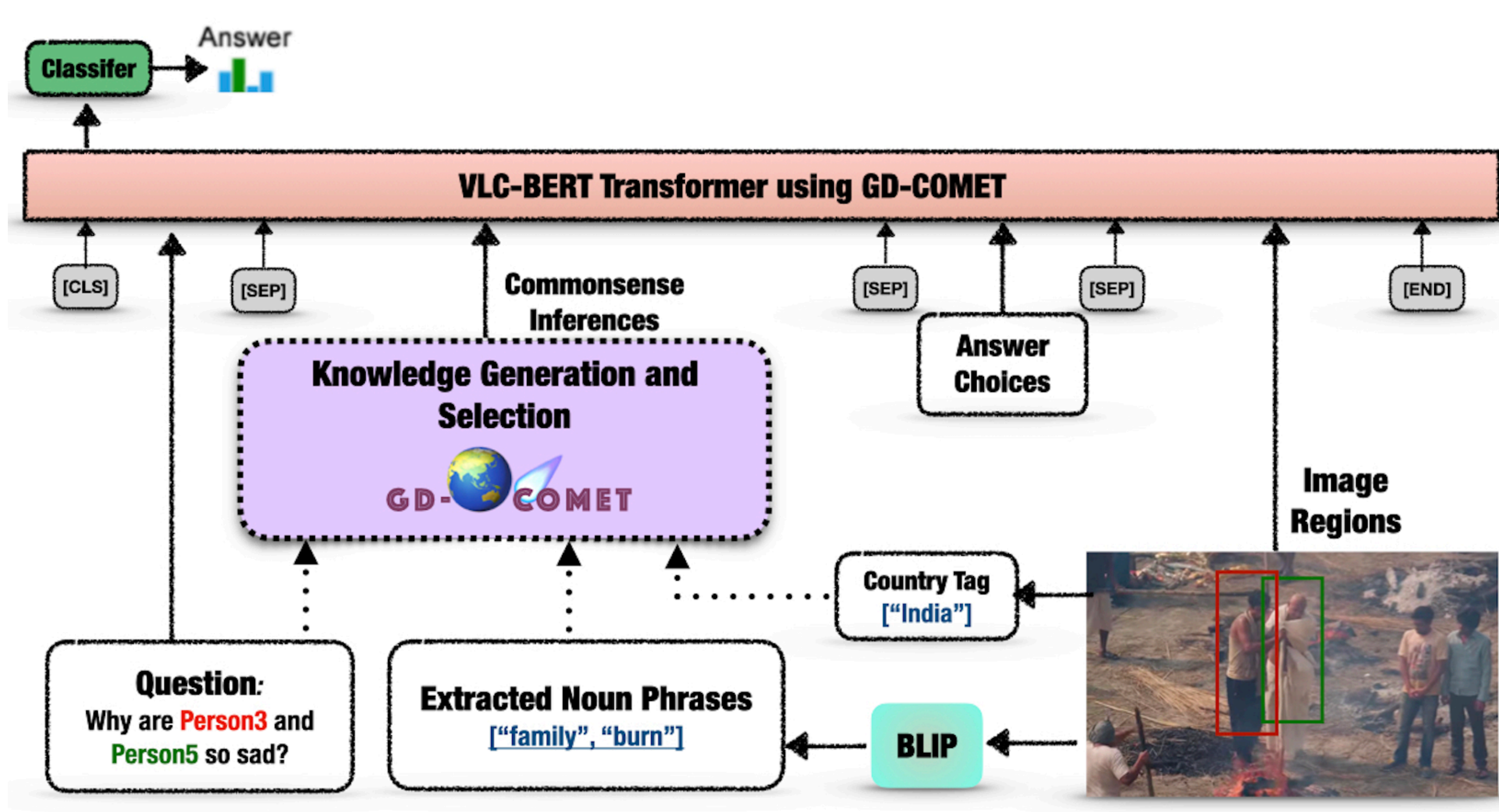


VLC-BERT: Visual Question Answering with Contextualized Commonsense Knowledge. Sahithya Ravi, Aditya Chinchure, Leonid Sigal, Renjie Liao, and Vered Shwartz. WACV 2023.

GD-COMET: A Geo-Diverse Commonsense Inference Model. Mehar Bhatia and Vered Shwartz. EMNLP 2023.



# Extrinsic Evaluation



VLC-BERT: Visual Question Answering with Contextualized Commonsense Knowledge. Sahithya Ravi, Aditya Chinchure, Leonid Sigal, Renjie Liao, and Vered Shwartz. WACV 2023.

GD-COMET: A Geo-Diverse Commonsense Inference Model. Mehar Bhatia and Vered Shwartz. EMNLP 2023.

# From Cultural Concepts to Cultural Norms

## **Narrative**

Mike and Emily had a delightful dinner in a restaurant in Madrid. The bill was 96 Euros, so they put a 100 Euro bill on the table and left the restaurant.

# From Cultural Concepts to Cultural Norms

## Narrative

Mike and Emily had a delightful dinner in a restaurant in Madrid. The bill was 96 Euros, so they put a 100 Euro bill on the table and left the restaurant.

## Situation

Tipping 4% at a restaurant



# From Cultural Concepts to Cultural Norms

## Narrative

Mike and Emily had a delightful dinner in a restaurant in Madrid. The bill was 96 Euros, so they put a 100 Euro bill on the table and left the restaurant.

## Situation

Tipping 4% at a restaurant

## Social Norms

 It's rude to tip less than 15%

 It's considered optional to tip

...

# In this talk...



## 1 Language Grounded in Non-linguistic Context



## 2 Reasoning



## 3 Real-World Applications



# Reasoning with Chain-of-Thought (CoT) Prompting

## Standard Prompting

### Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

### Model Output

A: The answer is 27. ❌

## Chain-of-Thought Prompting

### Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls.  $5 + 6 = 11$ . The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

### Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had  $23 - 20 = 3$ . They bought 6 more apples, so they have  $3 + 6 = 9$ . The answer is 9. ✅



# Reasoning with Chain-of-Thought (CoT) Prompting

## Standard Prompting

### Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

### Model Output

A: The answer is 27. ❌

## Chain-of-Thought Prompting

### Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls.  $5 + 6 = 11$ . The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

### Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had  $23 - 20 = 3$ . They bought 6 more apples, so they have  $3 + 6 = 9$ . The answer is 9. ✅

- Improved performance in few-shot and zero-shot reasoning tasks

# Reasoning with Chain-of-Thought (CoT) Prompting

## Standard Prompting

### Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

### Model Output

A: The answer is 27. ❌

## Chain-of-Thought Prompting

### Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls.  $5 + 6 = 11$ . The answer is 11.

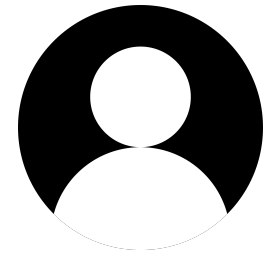
Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

### Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had  $23 - 20 = 3$ . They bought 6 more apples, so they have  $3 + 6 = 9$ . The answer is 9. ✅

- Improved performance in few-shot and zero-shot reasoning tasks
- Explainable predictions... but are they faithful?

# Example: CoT for Stance Detection



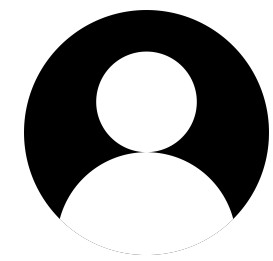
**Twitter User** @TwitterUser · Oct 13, 2016

She can't even manage her husband and  
she wants to be @POTUS

Is this tweet in favour or against Hillary Clinton?



# Example: CoT for Stance Detection



**Twitter User** @TwitterUser · Oct 13, 2016

She can't even manage her husband and  
she wants to be @POTUS

Is this tweet in favour or against Hillary Clinton?

```
# n in-context examples
```

```
tweet: <tweet1>
```

```
target: <stance target1>
```

```
Q: What is the tweet's stance on the target?
```

```
The options are: against, favor, none
```

```
Premise: <premise1>
```

```
Conclusion: <conclusion1>
```

```
...
```

```
# Test example
```

```
tweet: She can't even manage her husband and  
she wants to be @POTUS
```

```
target: Hillary Clinton
```

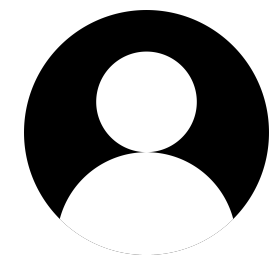
```
Q: What is the tweet's stance on the target?
```

```
The options are: against, favor, none
```

```
Premise: Hillary Clinton is not qualified to be  
president because of her bad managing abilities.
```

```
Conclusion: The author is against Hillary Clinton.
```

# Example: CoT for Stance Detection

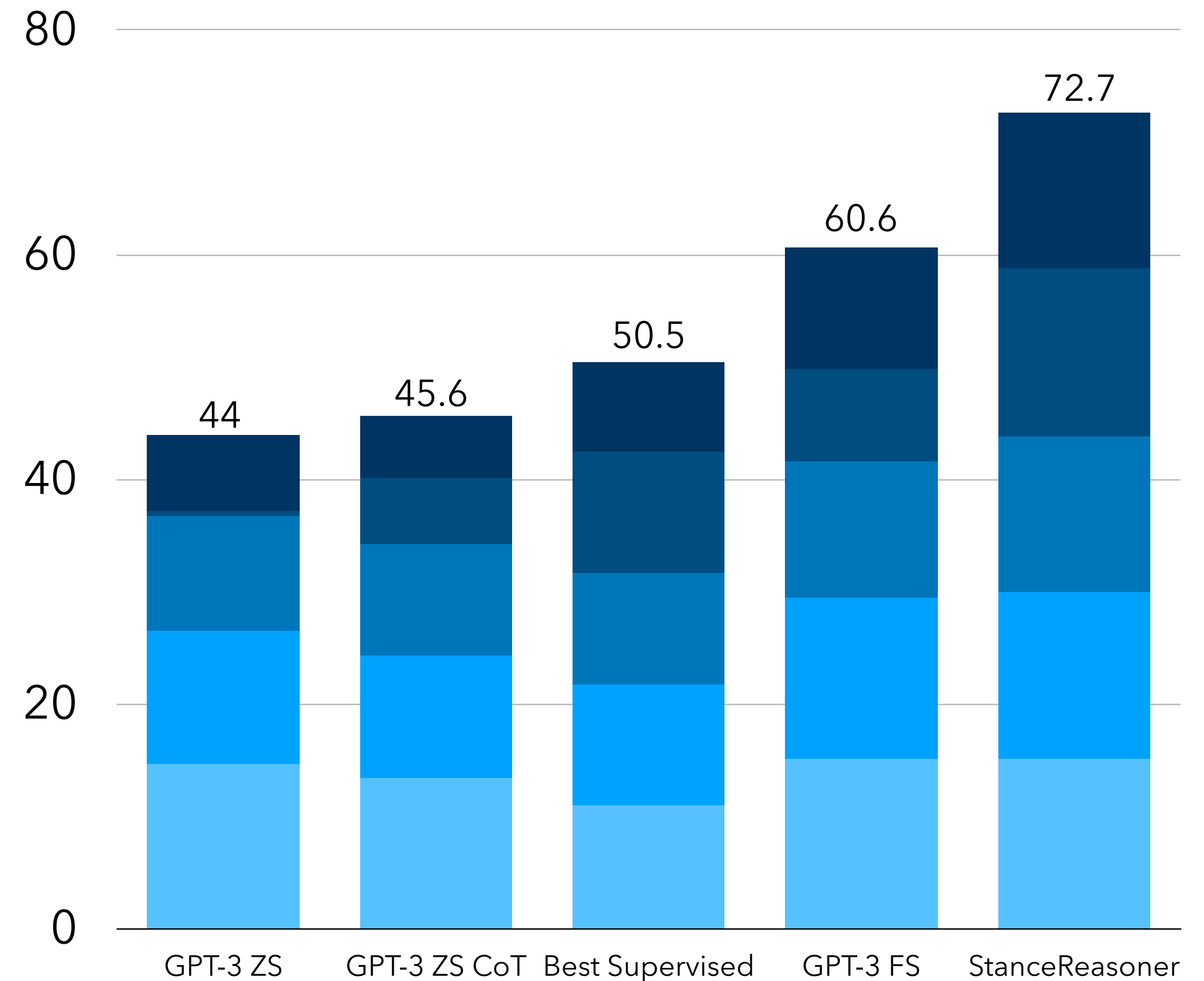


**Twitter User** @TwitterUser · Oct 13, 2016

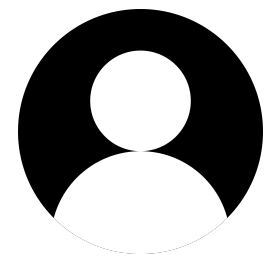
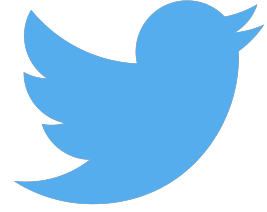
She can't even manage her husband and she wants to be @POTUS

Is this tweet in favour or against Hillary Clinton?

```
# n in-context examples
tweet: <tweet1>
target: <stance target1>
Q: What is the tweet's stance on the target?
The options are: against, favor, none
Premise: <premise1>
Conclusion: <conclusion1>
...
# Test example
tweet: She can't even manage her husband and
she wants to be @POTUS
target: Hillary Clinton
Q: What is the tweet's stance on the target?
The options are: against, favor, none
Premise: Hillary Clinton is not qualified to be
president because of her bad managing abilities.
Conclusion: The author is against Hillary Clinton.
```



# Beyond Deductive Reasoning...



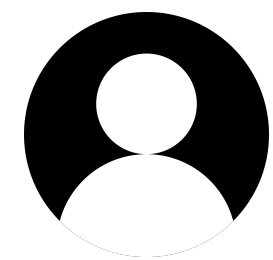
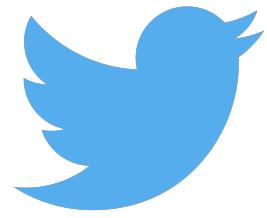
**Twitter User** @TwitterUser · Apr 13

When women spend too much time out of the kitchen  
they get over opinionated and think they know  
everything #feminist

Is this tweet in favour or against the feminist movement?



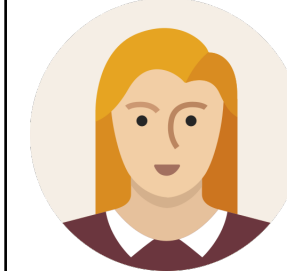
# Beyond Deductive Reasoning...



**Twitter User** @TwitterUser · Apr 13

When women spend too much time out of the kitchen they get over opinionated and think they know everything #feminist

Is this tweet in favour or against the feminist movement?

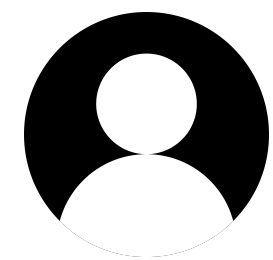
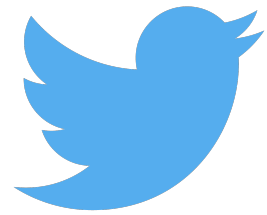


**Jane Doe** @FeministWoman · Apr 13

When women spend too much time out of the kitchen they get over opinionated and think they know everything #feminist

What about this one?

# Beyond Deductive Reasoning...



**Twitter User** @TwitterUser · Apr 13

When women spend too much time out of the kitchen they get over opinionated and think they know everything #feminist

Is this tweet in favour or against the feminist movement?



**Jane Doe** @FeministWoman · Apr 13

When women spend too much time out of the kitchen they get over opinionated and think they know everything #feminist

What about this one?

The tweet seems to be against the feminist movement. However, there is not enough information to determine this with certainty. For example, if this tweet was written by a known feminist user, they might be using sarcasm to respond to someone else's misogynistic tweet. In that case, the user is actually in favour of the feminist movement.

# Defeasible Inference in Natural Language

Given a premise  $P$  and hypothesis  $H$ , an update  $U$  is a:  
**Weakener** if a human would most likely find  $H$  *less likely to be true* after learning  $U$ ; and  
**Strengtheners** if they would find  $H$  *more likely to be true*.



# Defeasible Inference in Natural Language

Given a premise P and hypothesis H, an update U is a:  
**Weakener** if a human would most likely find H *less likely to be true* after learning U; and  
**Strengtheners** if they would find H *more likely to be true*.

## NLI

P: A group of people sitting around a table with laptops.

H: They have a work meeting.

S: They are in a conference room.

W: They are in a library.



# Defeasible Inference in Natural Language

Given a premise P and hypothesis H, an update U is a:  
**Weakener** if a human would most likely find H *less likely to be true* after learning U; and  
**Strengthenener** if they would find H *more likely to be true*.

## Commonsense

P: PersonX has a pool party.

H: Because PersonX wanted to hangout with friends.

S: It was PersonX's birthday.

W: PersonX was having a family reunion.



## NLI

P: A group of people sitting around a table with laptops.

H: They have a work meeting.

S: They are in a conference room.

W: They are in a library.





# Defeasible Inference in Natural Language

Given a premise P and hypothesis H, an update U is a:  
**Weakener** if a human would most likely find H *less likely to be true* after learning U; and  
**Strengthenener** if they would find H *more likely to be true*.

## Commonsense

P: PersonX has a pool party.

H: Because PersonX wanted to hangout with friends.

S: It was PersonX's birthday.

W: PersonX was having a family reunion.



## NLI

P: A group of people sitting around a table with laptops.

H: They have a work meeting.

S: They are in a conference room.

W: They are in a library.



## Social Norms

H: It is rude to comment on someone's weight.

S: They are sensitive to it.

W: You are a nutritionist.





# Defeasible Inference for Real-time Summarization

---

# Defeasible Inference for Real-time Summarization

Dec 2019

Mystery pneumonia outbreak in China sparks fear of deadly SARS virus

Mystery pneumonia outbreak in China sparks fear of deadly SARS virus ... The World Health Organization (WHO) criticised China for under reporting the number...



# Defeasible Inference for Real-time Summarization

## Dec 2019

Mystery pneumonia outbreak in China sparks fear of deadly SARS virus

Mystery pneumonia outbreak in China sparks fear of deadly SARS virus ... The World Health Organization (WHO) criticised China for under reporting the number...

## Jan 2020



China confirms human-to-human transmission of new coronavirus

20, a total of 224 cases of the new coronavirus infection were reported in China, including 217 confirmed cases: Five in Beijing. 14 in Guangdong. 198 in

China locks down city at center of virus outbreak

The World Health Organization has delayed a decision on declaring a global public health emergency. People wear face masks as they wait at a...



# Defeasible Inference for Real-time Summarization

## Dec 2019

Mystery pneumonia outbreak in China sparks fear of deadly SARS virus

Mystery pneumonia outbreak in China sparks fear of deadly SARS virus ... The World Health Organization (WHO) criticised China for under reporting the number...



## Feb 2020

The US Has Its First 'Community Spread' Coronavirus Case

The US Has Its First 'Community Spread' Coronavirus Case ... The patient was admitted to UC Davis Medical Center on February 19 but not diagnosed until

## Jan 2020



China confirms human-to-human transmission of new coronavirus

20, a total of 224 cases of the new coronavirus infection were reported in China, including 217 confirmed cases: Five in Beijing. 14 in Guangdong. 198 in

China locks down city at center of virus outbreak

The World Health Organization has delayed a decision on declaring a global public health emergency. People wear face masks as they wait at a...

# Defeasible Inference for Real-time Summarization

## Dec 2019

Mystery pneumonia outbreak in China sparks fear of deadly SARS virus

Mystery pneumonia outbreak in China sparks fear of deadly SARS virus ... The World Health Organization (WHO) criticised China for under reporting the number...



## Feb 2020

The US Has Its First 'Community Spread' Coronavirus Case

The US Has Its First 'Community Spread' Coronavirus Case ... The patient was admitted to UC Davis Medical Center on February 19 but not diagnosed until

## Jan 2020



China confirms human-to-human transmission of new coronavirus

20, a total of 224 cases of the new coronavirus infection were reported in China, including 217 confirmed cases: Five in Beijing. 14 in Guangdong. 198 in

China locks down city at center of virus outbreak

The World Health Organization has delayed a decision on declaring a global public health emergency. People wear face masks as they wait at a...

## Mar 2020



The WHO Just Declared Coronavirus COVID-19 a Pandemic

The World Health Organization (WHO) on March 11 declared COVID-19 a pandemic, pointing to the over 118,000 cases of the coronavirus illness in over 110...

# Defeasible Inference for Real-time Summarization

## Dec 2019

Mystery pneumonia outbreak in China sparks fear of deadly SARS virus

Mystery pneumonia outbreak in China sparks fear of deadly SARS virus ... The World Health Organization (WHO) criticised China for under reporting the number...



## Feb 2020

The US Has Its First 'Community Spread' Coronavirus Case

The US Has Its First 'Community Spread' Coronavirus Case ... The patient was admitted to UC Davis Medical Center on February 19 but not diagnosed until

## Jan 2020



China confirms human-to-human transmission of new coronavirus

20, a total of 224 cases of the new coronavirus infection were reported in China, including 217 confirmed cases: Five in Beijing. 14 in Guangdong. 198 in

China locks down city at center of virus outbreak

The World Health Organization has delayed a decision on declaring a global public health emergency. People wear face masks as they wait at a...



## Mar 2020

The WHO Just Declared Coronavirus COVID-19 a Pandemic

The World Health Organization (WHO) on March 11 declared COVID-19 a pandemic, pointing to the over 118,000 cases of the coronavirus illness in over 110...

...



# In this talk...



## 1 Language Grounded in Non-linguistic Context



## 2 Reasoning



## 3 Real-World Applications



# **LLMs hold both Promises and Risks for Real-World Applications**

# LLMs hold both Promises and Risks for Real-World Applications

Medical chatbot using  
OpenAI's GPT-3 told a  
fake patient to kill  
themselves



# LLMs hold both Promises and Risks for Real-World Applications

Medical chatbot using  
OpenAI's GPT-3 told a  
fake patient to kill  
themselves

## **OpenAI faces Canadian privacy probe alleging personal info harvesting**

Canada's head privacy commissioner is investigating the AI firm, and watchdogs in Germany, France, Ireland and Spain could be mulling similar action.

# LLMs hold both Promises and Risks for Real-World Applications

Medical chatbot using OpenAI's GPT-3 told a fake patient to kill themselves

## **OpenAI faces Canadian privacy probe alleging personal info harvesting**

Canada's head privacy commissioner is investigating the AI firm, and watchdogs in Germany, France, Ireland and Spain could be mulling similar action.

## **Australian mayor prepares world's first defamation lawsuit over ChatGPT content**

**ChatGPT falsely identified Brian Hood as guilty party in foreign bribery scandal. In reality he blew the whistle on the illegal scheme**

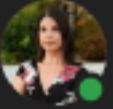
Who would you go to if you have a problem with your ear,  
a doctor that read all the medical literature or an ENT?



Who would you go to if you have a problem with your ear,  
a doctor that read all the medical literature or an ENT?

Who would you go to if you're getting divorced,  
a person who knows all the laws in all countries or a family lawyer in your province?



# Developing NLP Systems for Highly-specific Domains

**Vered Shwartz**  
February 7 · 🌐




Any guidelines on whether to submit a CTA claim or go to small claims court? And anything I should know before doing this?


My flight with Air Canada in December was delayed by 4 hours, which caused me to lose my connecting flight, and I was stranded in Toronto for 2 days on my way to the final destination. I had to pay the hotel expenses on my own. I filed a claim with AC but they got back to me today saying it was not within their control. I don't think it's true and I believe they are just trying to get out of compensating us (I can elaborate more).

Thanks in advance!

  7


16 comments

 Like  Comment  Share

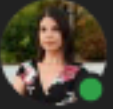
**Gabor Lukacs** [Admin](#)

1) I would personally avoid the CTA, and take the case to the Civil Resolutions Tribunal in BC instead.

2) Before doing that, I would consider sending Air Canada another email or letter, asking them to provide details of the events on the basis of which they claim it was outside their control. If they ignore it or refuse, it can be used as further evidence against Air Canada, and I would also seek compensation for breach of s. 19(4) of the APPR.

[Like](#) [Reply](#) [Share](#) 6w  6



# Developing NLP Systems for Highly-specific Domains

**Vered Shwartz**  
February 7 · 🌐




Any guidelines on whether to submit a CTA claim or go to small claims court? And anything I should know before doing this?


My flight with Air Canada in December was delayed by 4 hours, which caused me to lose my connecting flight, and I was stranded in Toronto for 2 days on my way to the final destination. I had to pay the hotel expenses on my own. I filed a claim with AC but they got back to me today saying it was not within their control. I don't think it's true and I believe they are just trying to get out of compensating us (I can elaborate more).

Thanks in advance!

  7


16 comments

 Like  Comment  Share

**Gabor Lukacs** Admin

1) I would personally avoid the CTA, and take the case to the Civil Resolutions Tribunal in BC instead.


2) Before doing that, I would consider sending Air Canada another email or letter, asking them to provide details of the events on the basis of which they claim it was outside their control. If they ignore it or refuse, it can be used as further evidence against Air Canada, and I would also seek compensation for breach of s. 19(4) of the APPR.

Like Reply Share 6w  6

VE Any guidelines on whether to submit a CTA claim or go to small claims court? And anything I should know before doing this?

My flight with Air Canada in December was delayed by 4 hours, which caused me to miss my connecting flight, and I was stranded in Toronto for 2 days on my way to the final destination. I had to pay the hotel expenses on my own. I filed a claim with AC but they got back to me today saying it was not within their control. I don't think it's true and I believe they are just trying to get out of compensating us (I can elaborate more).

Thanks in advance!

 I'm sorry to hear about your experience with Air Canada. If you feel that your claim has been wrongly denied, there are a few options available to you.

Firstly, you could consider submitting a claim under the Canadian Transportation Agency's (CTA) Air Passenger Protection Regulations. These regulations provide for compensation for flight delays, cancellations, and other issues that may affect air travel. You can find more information about the process on the CTA's website.

Alternatively, you could file a claim in small claims court. Small claims court is designed to handle relatively minor disputes between individuals and businesses, and it can be a relatively simple and inexpensive way to pursue a claim. However, you may need to consult with a lawyer or seek legal advice before pursuing this option.

Before taking any action, it's important to gather all of the relevant information and documentation, including your flight itinerary, receipts for any expenses you incurred, and any correspondence with Air Canada. You may also want to consider reaching out to a consumer rights organization or seeking legal advice to ensure that you understand your rights and the best course of action.

Good luck with your claim!



# Chatbot for Air Passenger Rights

# Chatbot for Air Passenger Rights

- Goal: reduce the workload of the human volunteers

# Chatbot for Air Passenger Rights

- Goal: reduce the workload of the human volunteers
- High precision (**no hallucinations!**),  
low recall is tolerated



# Chatbot for Air Passenger Rights

Air Passenger Rights Chatbot ✈️

I need help with a canceled flight and damaged luggage. Can I get on another flight without paying more, and what should I do about the luggage?

🔍 Results

What are the policies for getting on another flight after a flight cancellation?

**Flight Cancellation: Departing from the EU or EEA**

Your Options: Continue or Cancel Trip The airline must offer you the choice between continuing or cancelling your trip. If you choose to continue your travel, the airline must offer you the choice between: transportation to your destination at the earliest opportunity; or transportation to your destination at a later date, subject to availability of seats. If you choose to cancel your trip, the airline must transport you to the first point of departure at the earliest opportunity, and refund your ticket.

What steps should be taken when luggage is damaged during a flight?

**Damaged Baggage: Step by Step Guide**

Inspect your baggage As soon as you take it off the carousel, inspect your baggage. Document any and all damage to your baggage Take photos. Record a video with your phone. Report the damage at the airport

👍

👎

Ask your question or describe your issue... 

↑

Empowering Air Travelers: A Chatbot for Canadian Air Passenger Rights.  
Maksym Taranukhin, Sahithya Ravi, Gábor Lukács, Evangelos Milios and Vered Shwartz. arXiv 2024.

# Chatbot for Air Passenger Rights

- Goal: reduce the workload of the human volunteers
- High precision (**no hallucinations!**), low recall is tolerated

Air Passenger Rights Chatbot ✈️

I need help with a canceled flight and damaged luggage. Can I get on another flight without paying more, and what should I do about the luggage?

🔍 Results

What are the policies for getting on another flight after a flight cancellation?

**Flight Cancellation: Departing from the EU or EEA**

Your Options: Continue or Cancel Trip The airline must offer you the choice between continuing or cancelling your trip. If you choose to continue your travel, the airline must offer you the choice between: transportation to your destination at the earliest opportunity; or transportation to your destination at a later date, subject to availability of seats. If you choose to cancel your trip, the airline must transport you to the first point of departure at the earliest opportunity, and refund your ticket.


What steps should be taken when luggage is damaged during a flight?

**Damaged Baggage: Step by Step Guide**

Inspect your baggage As soon as you take it off the carousel, inspect your baggage. Document any and all damage to your baggage Take photos. Record a video with your phone. Report the damage at the airport

👍

👎

Ask your question or describe your issue... 



# Chatbot for Air Passenger Rights

- Goal: reduce the workload of the human volunteers
- High precision (**no hallucinations!**), low recall is tolerated
- Maintain user privacy

Air Passenger Rights Chatbot ✈️

I need help with a canceled flight and damaged luggage. Can I get on another flight without paying more, and what should I do about the luggage?

🔍 Results

What are the policies for getting on another flight after a flight cancellation?

**Flight Cancellation: Departing from the EU or EEA**

Your Options: Continue or Cancel Trip The airline must offer you the choice between continuing or cancelling your trip. If you choose to continue your travel, the airline must offer you the choice between: transportation to your destination at the earliest opportunity; or transportation to your destination at a later date, subject to availability of seats. If you choose to cancel your trip, the airline must transport you to the first point of departure at the earliest opportunity, and refund your ticket.


What steps should be taken when luggage is damaged during a flight?

**Damaged Baggage: Step by Step Guide**

Inspect your baggage As soon as you take it off the carousel, inspect your baggage. Document any and all damage to your baggage Take photos. Record a video with your phone. Report the damage at the airport

👍

👎

Ask your question or describe your issue... 

# Recap

 LLMs enable exciting new research directions in NLP.

 Remaining challenges and future directions include:

 Reasoning about extra-linguistic contexts

 Advanced reasoning beyond deductive reasoning (defeasible, abductive, counterfactual...)

 NLP for real-world applications

 Evaluation remains an issue... we need it for better tracking the progress!



# Recap

 LLMs enable exciting new research directions in NLP.

 Remaining challenges and future directions include:

 Reasoning about extra-linguistic contexts

 Advanced reasoning beyond deductive reasoning (defeasible, abductive, counterfactual...)

 NLP for real-world applications

 Evaluation remains an issue... we need it for better tracking the progress!

## Thank You!

 @VeredShwartz

 vshwartz@cs.ubc.ca