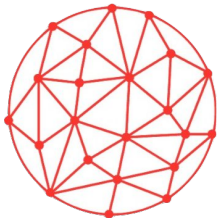


# Can LLMs Assist Annotators in Identifying Morality Frames? - Case Study on Vaccination Debate on Social Media

**Tunazzina Islam**, Dan Goldwasser

Department of Computer Science,

Purdue University, West Lafayette, IN 47907, USA



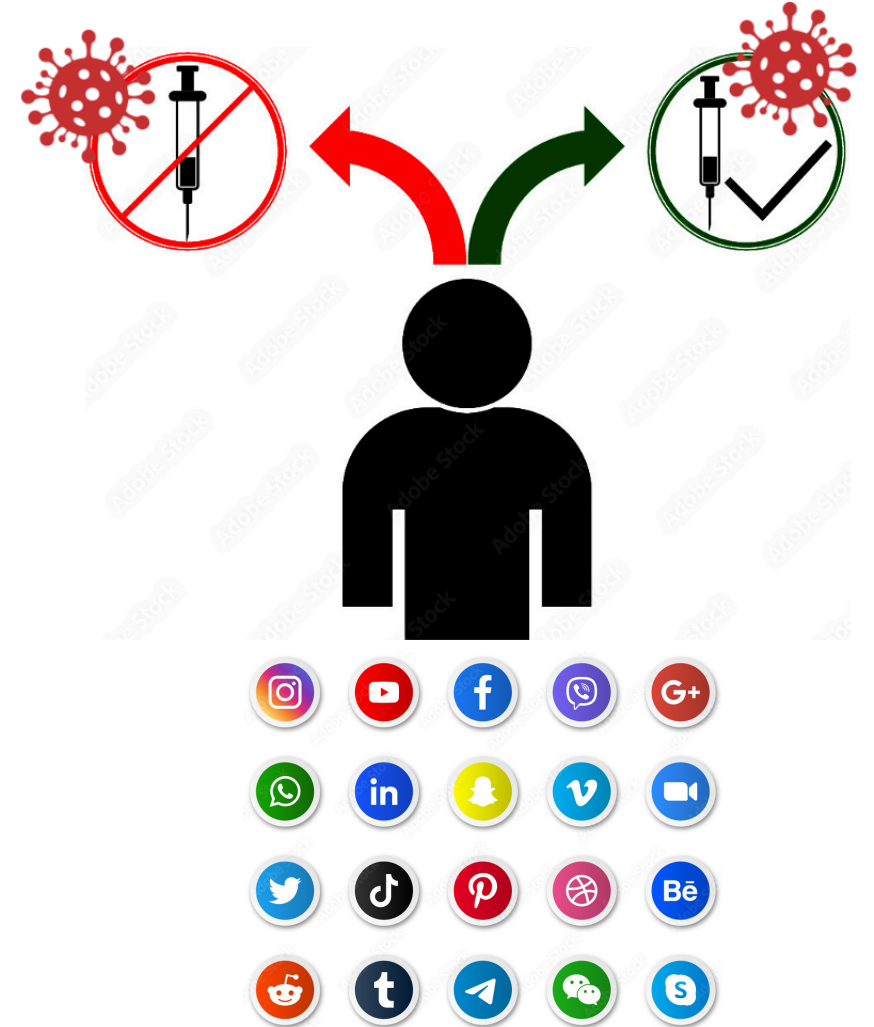
**ACM Web Science 2025**

May 20-23, 2025 - New Brunswick, NJ



# *Polarization on Social Media: Vaccination*

- Amplified public discourse on polarized topics (e.g., vaccinations).
- Disseminated both crucial *information* and *misinformation*.
- Impacted public health efforts through the “*infodemic*”.



# *Moral Perspective* of Vaccination

## The Role of Morality in Vaccination Debates

- Vaccine debates are shaped by **diverse moral perspectives**.
- Essential to build **effective communication and intervention tools**.



# Moral Perspective of Vaccination

## 🕒 The Role of Morality in Vaccination Debates

- Vaccine debates are shaped by **diverse moral perspectives**.
- Essential to build **effective communication and intervention tools**.

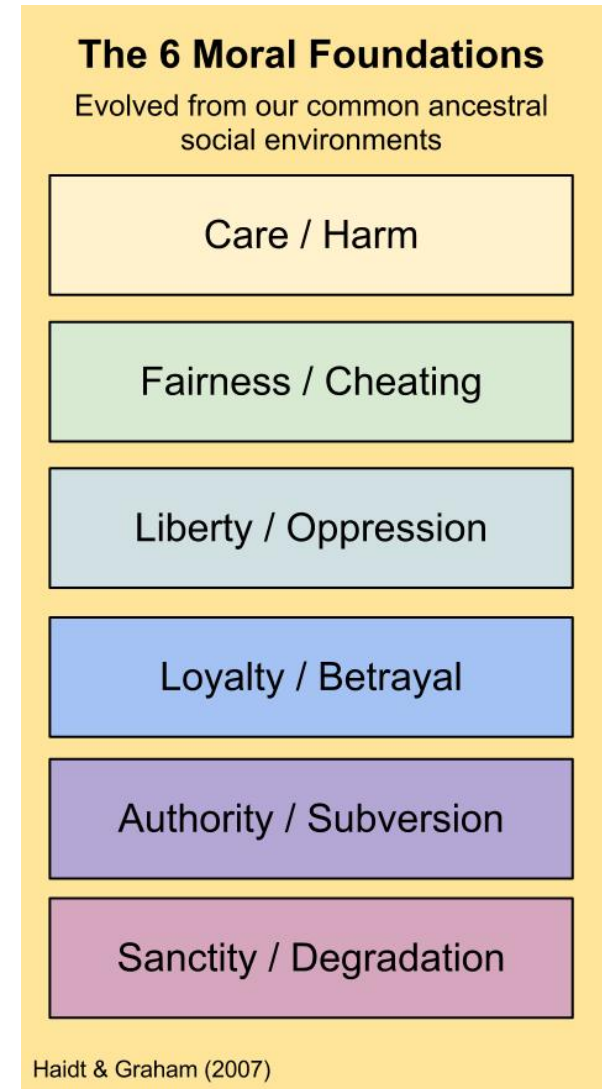
## 🧠 What is Morality Frame Analysis?

- A tool to uncover **moral beliefs** underlying opinions.
- Maps discussions to **moral foundations** (e.g., care/harm, liberty/oppression).
- Identifies **moral sentiments** toward **key entities/roles**.



# Moral Foundation Theory (MFT)

- Identifies **six** moral foundations, each with *positive* and *negative* polarity.
- Useful for analyzing **how people form moral judgments and sentiments**.





# Moral Foundation Theory (MFT)

## 1. Care/Harm

**Reflects the base of Maslow's Hierarchy of Needs  
(Security, Shelter, Food, Water, Warmth)**

## 2. Liberty/Oppression

**Physical and Mental Freedom  
Social Intolerance of Bullies**

## 3. Fairness/Cheating

**Equality of Opportunities  
Social Intolerance of "Free-Riders"**

# Moral Foundation Theory (MFT)

## 4. Loyalty/Betrayal

**Personal Trust, Group Identity, Patriotism**

**Social isolation of those who betray**

## 5. Authority/Subversion

**Competitive advantage of organized groups**

**Deference to “good” leaders (Alexander the Great)**

**Social intolerance of those who subvert the system**

## 6. Sanctity/Degradation

**Not simply a religious value**

**Respect for the human spirit**

**Social aversion of personal degradation**

Adapted from: Haidt, *The Righteous Mind*, Vintage Books (2012)

# Morality Frame Analysis

**Pfizer vaccine** testing utilized cell lines from human fetus tissue. This makes it abhorrent, heretical and blasphemous to anyone calling themselves a **Christian**.

## Reason Explanation:

use of fetal cell lines in vaccine development, framing this as a **violation of sacred values**, specifically within the Christian context.

## Morality Frame Analysis:

**Moral Foundation:** Degradation

**Negative Actors:** Pfizer Vaccine

**Negative Targets:** Christian



# ***Morality Framing: Annotation***

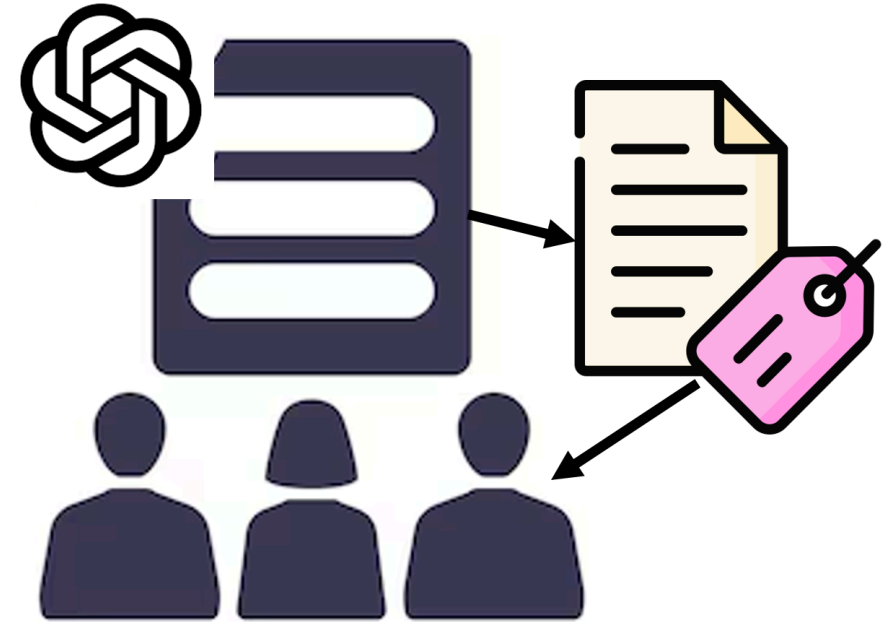
- Relying **solely** on **human annotators**
  - costly
  - Time-consuming
  - Inconsistency due to cognitive load

# ***Morality Framing: Annotation***

- Relying **solely** on **human annotators**
  - **costly**
  - **Time-consuming**
  - **Inconsistency due to cognitive load**
- Leveraging Large Language Models (LLMs)
  - Few-shot learning
    - ▶ In-context **examples** with **explanations**.

# *Human-AI Collaborative Annotation*

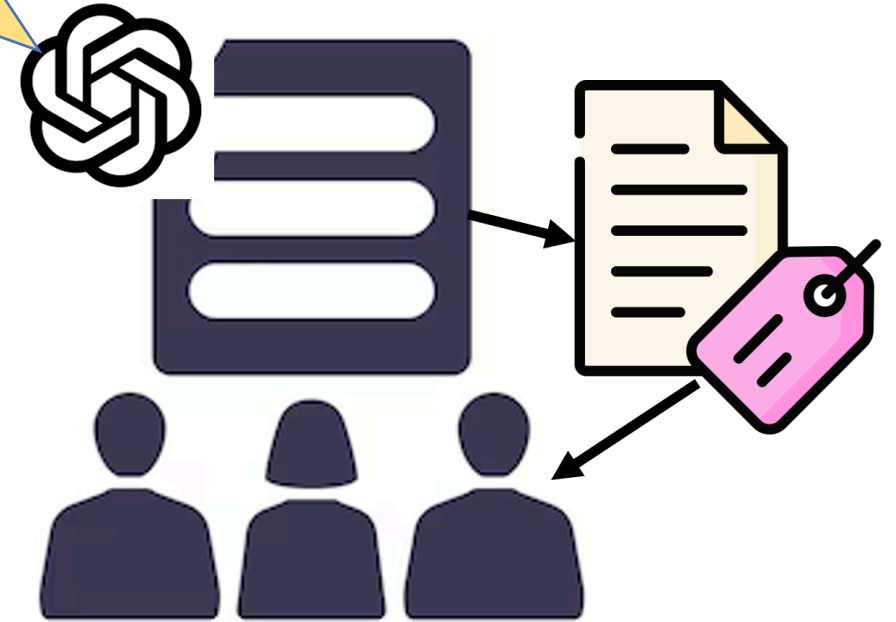
- **Two-step process:**
  - LLM-generated **concepts** + **explanations**.
- **Human evaluation**
  - ▶ Web-based annotation tool.



# Human-AI Collaborative Annotation

GPT-4o

- **Two-step process:**
  - LLM-generated **concepts** + **explanations**.
- **Human evaluation**
  - ▶ Web-based annotation tool.



# Morality Frame Identification: LLMs

## Instructions

## Definition of Moral Foundations

## Definition of Roles

## 7 examples with explanations

## Test Text

### Instructions:

1. Answer moral foundation of given text. If the text does not make any moral judgement, answer "none". For this case, you don't have to answer actor-target polarity.
2. If you answer a moral foundation other than "none", you need to answer actor-target polarity.
3. Answer the actor-target role with polarity of corresponding moral foundation. Answer who the Positive Actor / Positive Target / Negative Actor / Negative Target is.
4. Provide explanations of your answers.

For answering Moral Foundation, please look into the following categories:

1. care/harm: <definition>
2. fairness/cheating: <definition>
3. loyalty/betrayal: <definition>
4. authority/subversion: <definition>
5. sanctity/degradation: <definition>
6. liberty/oppression: <definition>
7. none: <definition>

For Answering actor and target roles with polarity, please look into following definitions: <definition>

Text: <Text1>

Question1: What's the moral foundation of the above text? Give the explanation of choosing the certain answer.

Answer1: <moral foundation1>

Explanation1: <moral foundation1 explanation>

Question2: Answer the actor-target role with polarity of corresponding moral foundation. Please only choose actor-target role with polarity from the given text. Don't generate a new actor-target role with polarity. Give the explanation of choosing the certain answer.

Answer2: <actor-target-polarity role1>

Explanation2: <role1 explanation>

.....  
.....

Text: <Test Text>

Question1: What's the moral foundation of the above text? Give the explanation of choosing the certain answer.

Answer1: <predicted moral foundation>

Explanation1: <moral foundation explanation>

Question2: Answer the actor-target role with polarity of corresponding moral foundation. Please only choose actor-target role with polarity from the given text. Don't generate a new actor-target role with polarity. Give the explanation of choosing the certain answer.

Answer2: <predicted actor-target-polarity role>

Explanation2: <role explanation>



# Web-based Annotation Tool: Task Interface

Text

LLM-generated MF  
with explanations

LLM-generated Roles  
with explanations

**Text:** *#BidenIsADisgrace can't back up his rhetoric regarding Mr. Trump's vaccine stances and mandate a #COVID19 vaccine. He'd rather cheerlead for businesses, schools and any other groups to do his work for him. Why can't we have a President that actually stands up?? #BidenisaFailure*

**Moral Foundation:** *authority/subversion*

**MF Explanation:** *because the text criticizes President Biden's actions and leadership, implying that he is not doing his job properly and should be taking a stronger stance*

**Actor-Target polarity role:** *positive actor: , negative actor: Biden, positive target: , negative target: Americans (implied)*

**Role Explanation:** *President Biden is the negative actor who is not taking proper action, and Americans are the negative targets who are affected by his inaction*

**Do you think the given moral foundation and actor-target polarity role of the provided text are correct?**

☐ yes ☐ no

After finishing the task, please click **Submit** button.

Submit

Showing Task 1 of 11

Next HIT

# Task Interface: Human Disagrees with LLMs

*Text:* When I see <sup>neg tar</sup>people rejecting the vaccine in the south, I don't only see Trump supporters or conspiracy nutters. What I see is a lot of fear and distrust, <sup>neg act</sup>past hurt and betrayal that must've inspired this heightened paranoia of a violent world and medical establishment.

**Moral Foundation:** care/harm

**MF Explanation:** because the author is expressing concern for people who are fearful and distrustful due to past hurts and betrayals

**Actor-Target polarity role:** positive actor: , negative actor: past hurt and betrayal, positive target: , negative target: People rejecting the vaccine/Trump supporters/conspiracy nutters

**Role Explanation:** Because the text implies that the past hurt and betrayal (Negative Actor) have caused fear, distrust, and paranoia (Negative effect) in the people rejecting the vaccine/Trump supporters/conspiracy nutters (Negative Target)

**Do you think the given moral foundation and actor-target polarity role of the provided text are correct?**

☐ yes ☒ no

If you feel that the given moral foundation is not correct, choose the correct moral foundation from following:

☐ care/harm ☐ fairness/cheating ☒ loyalty/betrayal ☐ authority/subversion ☐ sanctity/degradation ☐ liberty/oppression ☐ none

**First** pick the color.

**Second** highlight the text for actor-target role with polarity associated with corresponding moral foundation.

Positive Actor

Positive Target

Negative Actor

Negative Target

# Participants Results & Survey

- 9 participants
- 150 tweets (*pacheco et al., 2022*)
- Krippendorff's  $\alpha = 0.979$  (satisfactory level of agreement)

# Participants Results & Survey

- 9 participants
- 150 tweets (*pacheco et al., 2022*)
- Krippendorff's  $\alpha = 0.979$  (satisfactory level of agreement)
- **Annotator Survey Insights**
  - **Task Difficulty:** Rated 1-5, with and without LLM labels/explanations.
    - 1  $\rightarrow$  very easy, 2  $\rightarrow$  easy, 3  $\rightarrow$  okay, 4  $\rightarrow$  hard, and 5  $\rightarrow$  very hard.
  - **Explanation Effectiveness:** Assessing LLMs' helpfulness in understanding frames.
  - **Cognitive Load:** Evaluating whether explanations ease cognitive demands.
  - **Time Efficiency:** Measuring time for batch completion.
  - **Feedback:** Annotators' comments on explanation efficiency and task complexity.

# Participants Results & Survey

- ✅ If majority of annotators agree with LLM-generated label → counted as a ‘win’.
- ❌ If majority vote ‘no’ → use annotators’ moral foundation and role labels for evaluation.
- 📊 Compute overall accuracy for morality frame prediction.
- 🔍 Evaluate moral foundation prediction using accuracy and macro average F1 score.

Method	$Acc_{overall}$	$Acc_{MF}$	$F1_{MF}$
<code>few-shot<sub>wo/expl</sub></code>	64.81%	71.7%	72.58%
<b><code>few-shot<sub>w/expl</sub></code></b>	<b>90.79%</b>	<b>92.67%</b>	<b>93.51%</b>

Table 1: Morality Frame Identification Task.



# Participants Results & Survey

- ✅ If majority of annotators agree with LLM-generated label → counted as a ‘win’.
- ❌ If majority vote ‘no’ → use annotators’ moral foundation and role labels for evaluation.
- 📊 Compute overall accuracy for morality frame prediction.
- 🔍 Evaluate moral foundation prediction using accuracy and macro average F1 score.
- 💡 Few-shot prompting w/ explanation achieves **higher** accuracy than w/o explanation.

Method	$Acc_{overall}$	$Acc_{MF}$	$F1_{MF}$
<i>few-shot<sub>wo/expl</sub></i>	64.81%	71.7%	72.58%
<b><i>few-shot<sub>w/expl</sub></i></b>	<b>90.79%</b>	<b>92.67%</b>	<b>93.51%</b>

Table 1: Morality Frame Identification Task.

# Participants Results & Survey

- Labels and explanations provided by LLMs:
  - Reduces task difficulty.

Participant	Task Diff. (w/o expl.)	Task Diff. (w/ expl.)	Expl. Helpful?	Red. Load?	Cog.	Avg. Time/Batch (min)
P1	very hard	easy	yes	yes		30
P2	very hard	easy	yes	yes		45
P3	very hard	easy	yes	yes		33
P4	hard	easy	yes	yes		70
P5	very hard	easy	yes	yes		42
P6	very hard	easy	yes	yes		31
P7	okay	easy	yes	yes		48
P8	hard	okay	yes	yes		30
P9	hard	easy	yes	yes		90

Table 2: Survey Results on Morality Frame Identification Task.

# Participants Results & Survey

- Labels and explanations provided by LLMs:
  - Reduces task difficulty.
  - Helpful.
  - Lowers cognitive load.

Participant	Task Diff. (w/o expl.)	Task Diff. (w/ expl.)	Expl. Helpful?	Red. Load?	Cog.	Avg. Time/Batch (min)
P1	very hard	easy	yes	yes		30
P2	very hard	easy	yes	yes		45
P3	very hard	easy	yes	yes		33
P4	hard	easy	yes	yes		70
P5	very hard	easy	yes	yes		42
P6	very hard	easy	yes	yes		31
P7	okay	easy	yes	yes		48
P8	hard	okay	yes	yes		30
P9	hard	easy	yes	yes		90

Table 2: Survey Results on Morality Frame Identification Task.

# Participants Results & Survey

- Labels and explanations provided by LLMs:
  - Reduces task difficulty.
  - Helpful.
  - Lowers cognitive load.
  - Streamlines annotation process.

Participant	Task Diff. (w/o expl.)	Task Diff. (w/ expl.)	Expl. Helpful?	Red. Load?	Cog.	Avg. Time/Batch (min)
P1	very hard	easy	yes	yes		30
P2	very hard	easy	yes	yes		45
P3	very hard	easy	yes	yes		33
P4	hard	easy	yes	yes		70
P5	very hard	easy	yes	yes		42
P6	very hard	easy	yes	yes		31
P7	okay	easy	yes	yes		48
P8	hard	okay	yes	yes		30
P9	hard	easy	yes	yes		90

Table 2: Survey Results on Morality Frame Identification Task.

# Analysis of Morality Framing

- Correlation Heatmap of **Moral Foundations (MF)**, **Reasons**, and **Stances**.
- Interpret **reasons** as distributions over MFs and stances (and vice-versa).

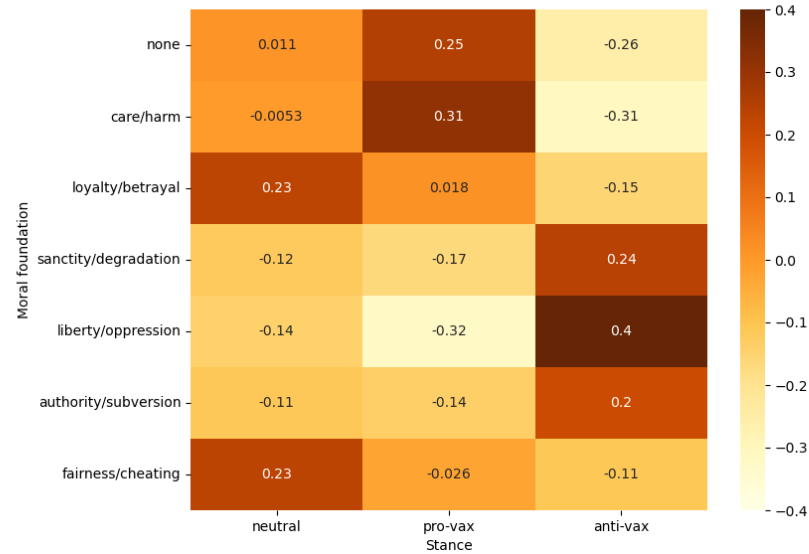


Fig. 1: MF & Stance

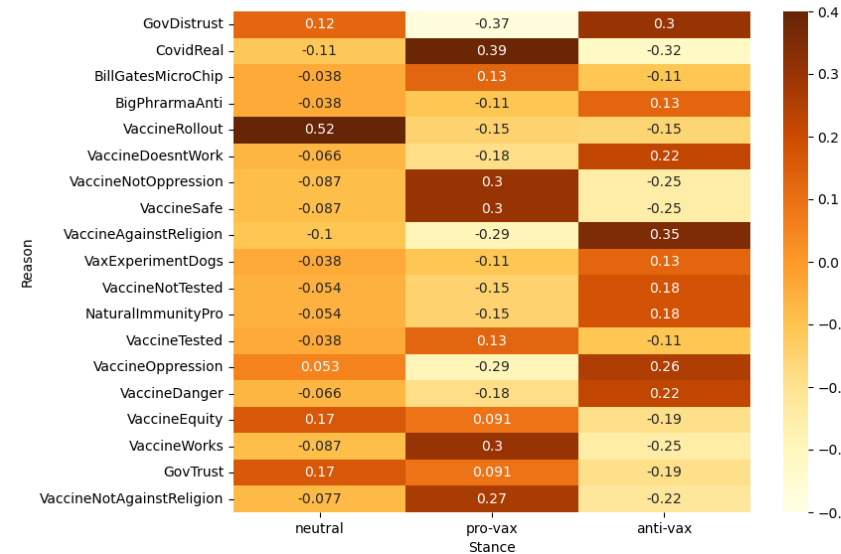


Fig. 2: Reasons & Stance

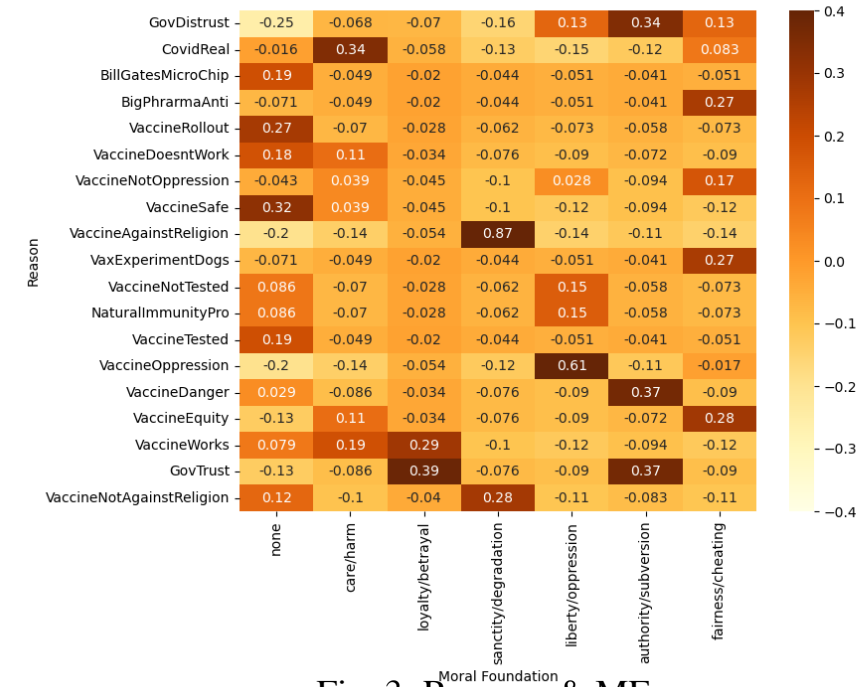







Fig. 3: Reasons & MF



# Analysis of Morality Framing

- Correlation Heatmap of **Moral Foundations (MF)**, **Reasons**, and **Stances**.
- Interpret **reasons** as distributions over MFs and stances (and vice-versa).

 Care/Harm → 'COVID is real' and 'vaccines work'  
 Liberty/Oppression → 'vaccine mandate is oppression'  
 Sanctity/Degradation → 'vaccine is (not) against religion'  
 Loyalty/Betrayal → 'trust in government'  
 Fairness/Cheating → 'vaccine equity'

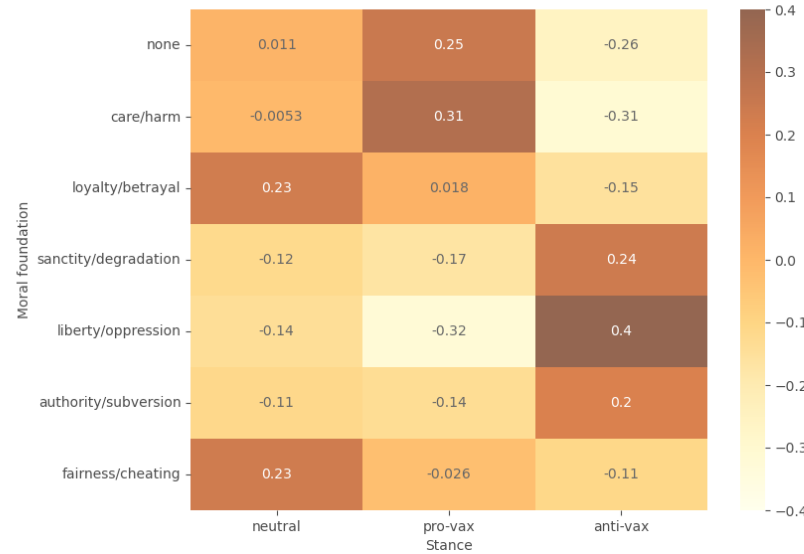


Fig. 1: MF & Stance

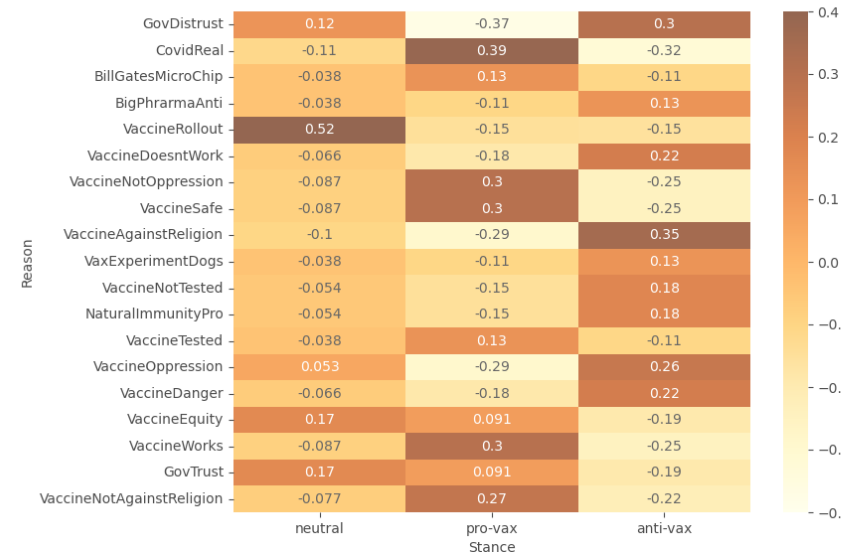


Fig. 2: Reasons & Stance

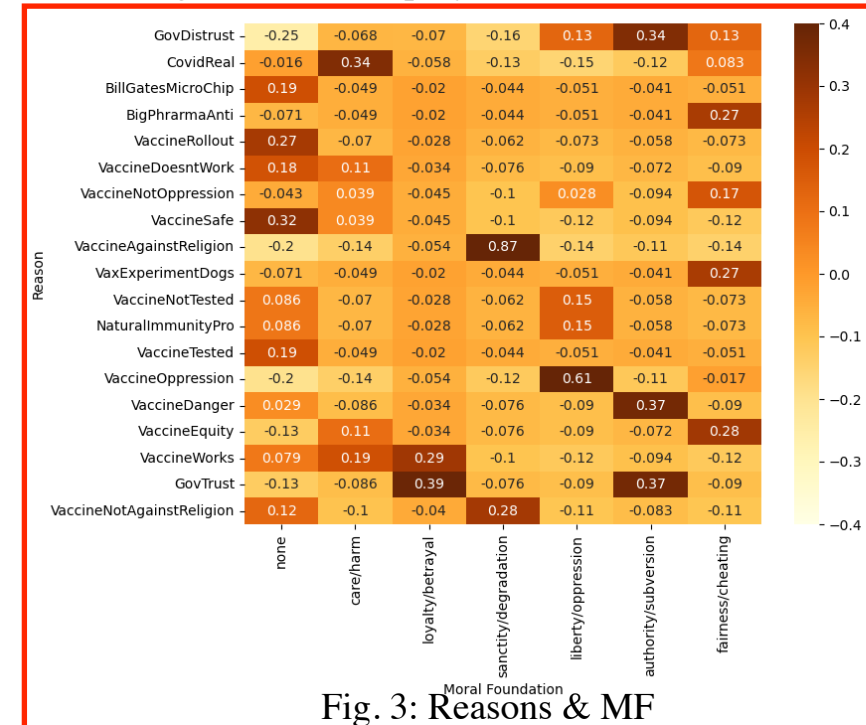











Fig. 3: Reasons & MF

# Analysis of Morality Framing

- Correlation Heatmap of **Moral Foundations (MF)**, **Reasons**, and **Stances**.
- Interpret **reasons** as distributions over MFs and stances (and vice-versa).

 Care/Harm → pro-vax   
 Liberty/Oppression → anti-vax 

 Care/Harm → 'COVID is real' and 'vaccines work'  
 Liberty/Oppression → 'vaccine mandate is oppression'  
 Sanctity/Degradation → 'vaccine is (not) against religion'  
 Loyalty/Betrayal → 'trust in government'  
 Fairness/Cheating → 'vaccine equity'

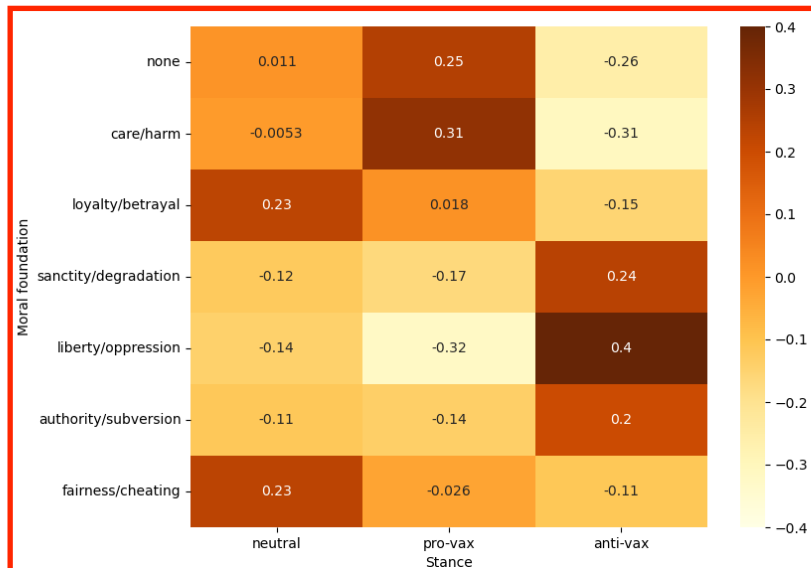


Fig. 1: MF & Stance



Fig. 2: Reasons & Stance

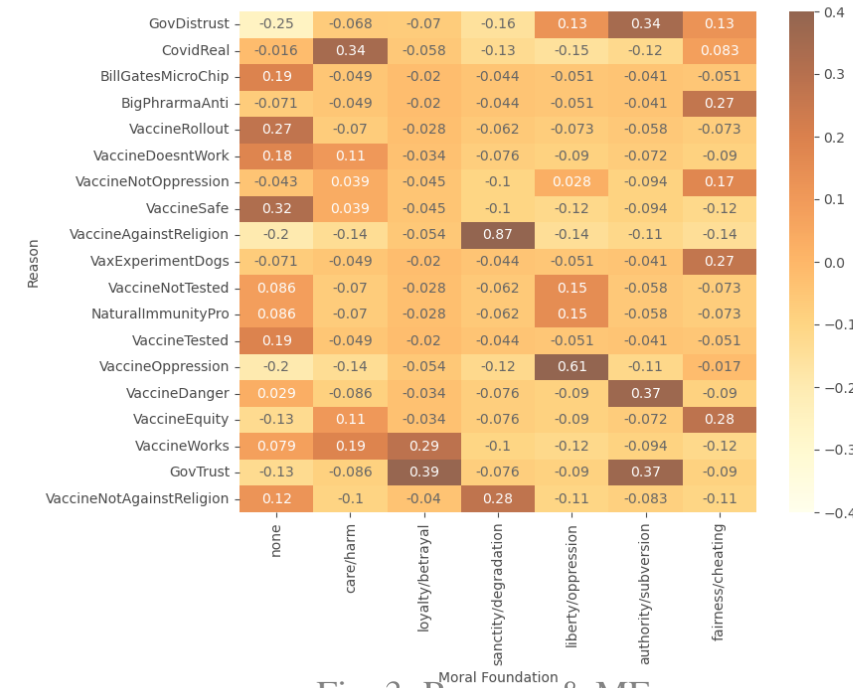








Fig. 3: Reasons & MF

# Analysis of Morality Framing

- Correlation Heatmap of **Moral Foundations (MF)**, **Reasons**, and **Stances**.
- Interpret **reasons** as distributions over MFs and stances (and vice-versa).

 Care/Harm → pro-vax   
 Liberty/Oppression → anti-vax 

 Pro-vax → 'COVID is real', 'vaccines work'  
 Anti-vax → 'vaccine is against religion', 'distrust in government', 'vaccine is oppression'






 Care/Harm → 'COVID is real' and 'vaccines work'  
 Liberty/Oppression → 'vaccine mandate is oppression'  
 Sanctity/Degradation → 'vaccine is (not) against religion'  
 Loyalty/Betrayal → 'trust in government'  
 Fairness/Cheating → 'vaccine equity'



Fig. 1: MF & Stance

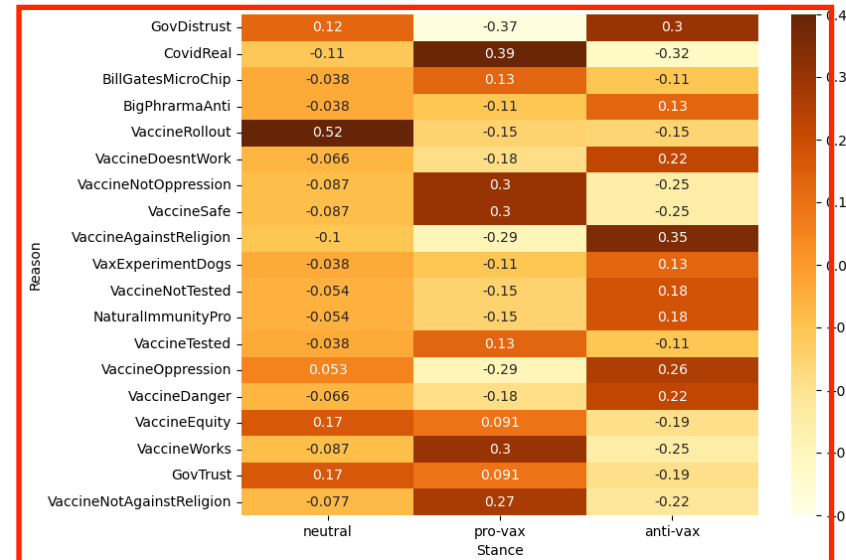


Fig. 2: Reasons & Stance

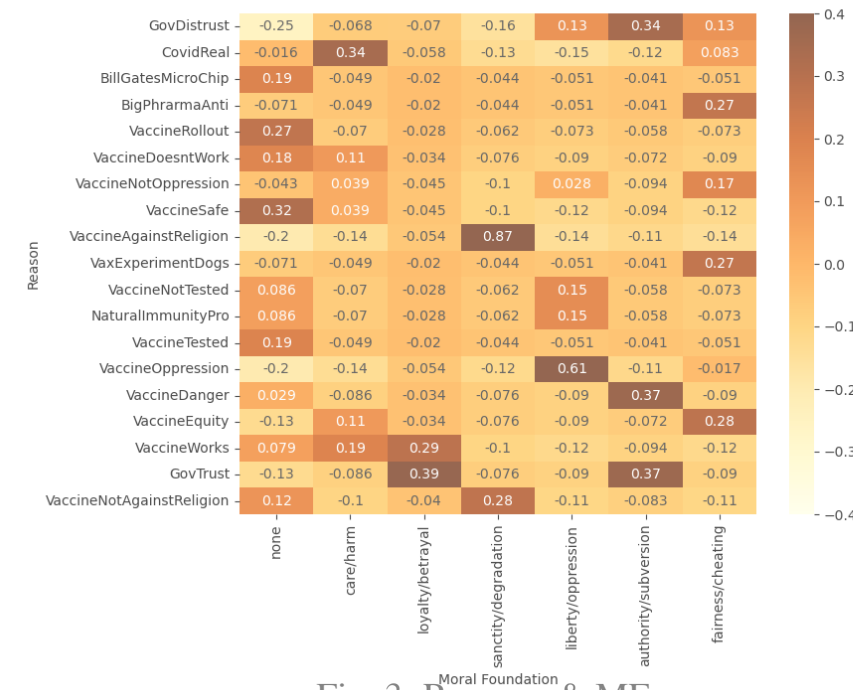


Fig. 3: Reasons & MF

# Analysis of Morality Framing

- Moral Foundations and Entity Roles with Polarity.
  - Roles: **Actors** and **Targets**
  - Polarity: **Positive** and **Negative**
- Helps reveal **how different perspectives construct moral arguments.**

# Analysis of Morality Framing

- Moral Foundations and Entity Roles with Polarity.
  - Roles: **Actors** and **Targets**
  - Polarity: **Positive** and **Negative**
- Helps reveal **how different perspectives construct moral arguments.**

MF	Top 2 Reasons for MF	Most Freq. Entity Roles
sanctity/degradation	VaccineAgainstReligion	100% anti-vax (vaccine, actor, negative) (Christian, target, negative)
	VaccineNotAgainstReligion	100% pro-vax (vaccine, actor, positive) (I, target, positive)
fairness/cheating	GovDistrust	67% anti-vax (Fauci, actor, negative) (children, target, negative)
	CovidReal	100% pro-vax (Fox News, actor, negative) (viewers, target, negative)
authority/subversion	GovDistrust	100% anti-vax (Biden, actor, negative) (people, target, negative)
	GovTrust	100% pro-vax (FDA, actor, positive) (Americans, target, positive)

Table 3: Top 2 reasons for sanctity/degradation, fairness/cheating, and authority/subversion moral foundations. Their most frequent opinions and entity roles.



# Analysis of Morality Framing



● Negative Actors: Fauci, Biden, Vaccines.

▼ Negative Targets: People, Children, Christians.

MF	Top 2 Reasons for MF	Most Freq. Entity Roles
sanctity/degradation	VaccineAgainstReligion	100% anti-vax (vaccine, actor, negative) (Christian, target, negative)
	VaccineNotAgainstReligion	100% pro-vax (vaccine, actor, positive) (I, target, positive)
fairness/cheating	GovDistrust	67% anti-vax (Fauci, actor, negative) (children, target, negative)
	CovidReal	100% pro-vax (Fox News, actor, negative) (viewers, target, negative)
authority/subversion	GovDistrust	100% anti-vax (Biden, actor, negative) (people, target, negative)
	GovTrust	100% pro-vax (FDA, actor, positive) (Americans, target, positive)

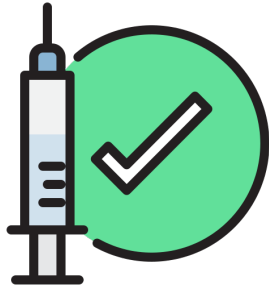
Table 3: Top 2 reasons for sanctity/degradation, fairness/cheating, and authority/subversion moral foundations. Their most frequent opinions and entity roles.

# Analysis of Morality Framing



● Negative Actors: Fauci, Biden, Vaccines.

▼ Negative Targets: People, Children, Christians.



● Negative Actors: Fox News.

▼ Negative Targets: Viewers.

● Positive Actors: FDA, vaccine.

◆ Positive Targets: Americans, I.

MF	Top 2 Reasons for MF	Most Freq. Entity Roles
sanctity/degradation	VaccineAgainstReligion	100% anti-vax (vaccine, actor, negative) (Christian, target, negative)
	VaccineNotAgainstReligion	100% pro-vax (vaccine, actor, positive) (I, target, positive)
fairness/cheating	GovDistrust	67% anti-vax (Fauci, actor, negative) (children, target, negative)
	CovidReal	100% pro-vax (Fox News, actor, negative) (viewers, target, negative)
authority/subversion	GovDistrust	100% anti-vax (Biden, actor, negative) (people, target, negative)
	GovTrust	100% pro-vax (FDA, actor, positive) (Americans, target, positive)

Table 3: Top 2 reasons for sanctity/degradation, fairness/cheating, and authority/subversion moral foundations. Their most frequent opinions and entity roles.

# Key Takeaways

- LLMs **support annotators** in complex psycholinguistic tasks like morality frame identification in vaccine debates.

# Key Takeaways

- LLMs **support annotators** in complex psycholinguistic tasks like morality frame identification in vaccine debates.
- Use **few-shot prompting with explanations** to guide LLMs in identifying:
  - Moral foundations
  - Actor-target polarity
- Integrate LLM-generated responses into the annotation process.

# Key Takeaways

- LLMs **support annotators** in complex psycholinguistic tasks like morality frame identification in vaccine debates.
- Use **few-shot prompting with explanations** to guide LLMs in identifying:
  - Moral foundations
  - Actor-target polarity
- Integrate LLM-generated responses into the annotation process.
- Integrating LLMs into the annotation process **enhances accuracy, reduces task difficulty, lowers cognitive load, streamlines the process.**

# Key Takeaways

- LLMs **support annotators** in complex psycholinguistic tasks like morality frame identification in vaccine debates.
- Use **few-shot prompting with explanations** to guide LLMs in identifying:
  - Moral foundations
  - Actor-target polarity
- Integrate LLM-generated responses into the annotation process.
- Integrating LLMs into the annotation process **enhances accuracy, reduces task difficulty, lowers cognitive load, streamlines the process.**
- Provide **qualitative analysis of morality frame with stances and reasons.**

# Key Takeaways

- LLMs **support annotators** in complex psycholinguistic tasks like morality frame identification in vaccine debates.
- Use **few-shot prompting with explanations** to guide LLMs in identifying:
  - Moral foundations
  - Actor-target polarity
- Integrate LLM-generated responses into the annotation process.
- Integrating LLMs into the annotation process **enhances accuracy, reduces task difficulty, lowers cognitive load, streamlines the process.**
- Provide **qualitative analysis of morality frame with stances and reasons.**
- Demonstrate a **collaborative human-AI framework** where LLMs augment—not replace—human judgment.

# Key Takeaways

- LLMs **support annotators** in complex psycholinguistic tasks like morality frame identification in vaccine debates.
- Use **few-shot prompting with explanations** to guide LLMs in identifying:
  - Moral foundations
  - Actor-target polarity
- Integrate LLM-generated responses into the annotation process.
- Integrating LLMs into the annotation process **enhances accuracy, reduces task difficulty, lowers cognitive load, streamlines the process.**
- Provide **qualitative analysis of morality frame with stances and reasons.**
- Demonstrate a **collaborative human-AI framework** where LLMs augment—not replace—human judgment.
- **Domain-agnostic Framework:** applicable to **political discourse or climate debates.**



# THANK YOU 😊

Slide: <https://tunazislam.github.io/files/AnnotMoralityLLMWebSci25.pdf>

## Questions?

Tunazzina Islam, Ph.D.

Department of Computer Science,  
Purdue University, West Lafayette, IN.

Email: [islam32@purdue.edu](mailto:islam32@purdue.edu)

 <https://tunazislam.github.io/>

 [@Tunaz\\_Islam](https://twitter.com/Tunaz_Islam)

