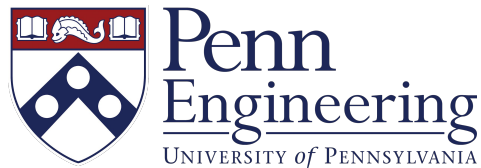


# Cultural Knowledge Injection

Shreya Havaladar, Lyle Ungar



# People talk to LLMs constantly



## Therapy

I've been feeling sad lately. Can I talk to you about it?

Of course. I'm here for you, and you can talk to me about anything — no judgment, no pressure.

Whenever you're ready, just start wherever it feels natural.

What's been weighing on you lately?

## Advice

My girlfriend is mad at me. What do I do?

I'm sorry to hear that — relationship tension can feel really heavy. Here's a way to start moving forward constructively:

## Education

Can you teach me how to impress my boss?

Definitely — impressing your boss isn't about grand gestures. It's about consistency, initiative, and emotional intelligence. Here's a breakdown of how to do it strategically:

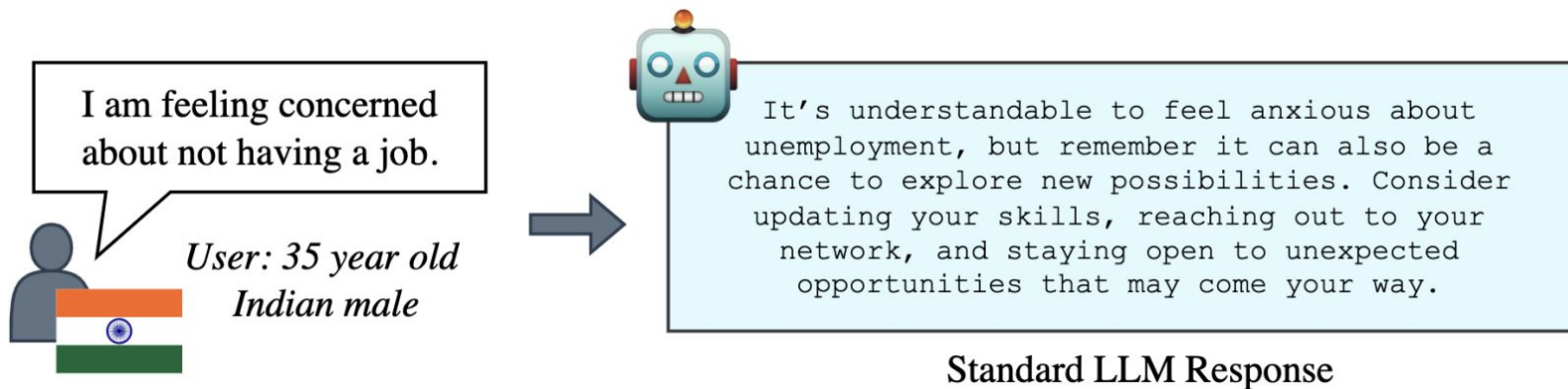
# But... conversational dynamics differ across cultures



Scenario with cultural variance	Stylistic axis that best captures this variance
Delivering bad news	Directness ↔ indirectness
Giving critical feedback	Politeness ↔ rudeness
Offering & accepting food or gifts	Insistence ↔ yielding
Asking for a favor	Gratitude ↔ expectation
Talking about personal failures & successes	Pride ↔ shame
Discussing sensitive topics	Self-disclosure ↔ secrecy

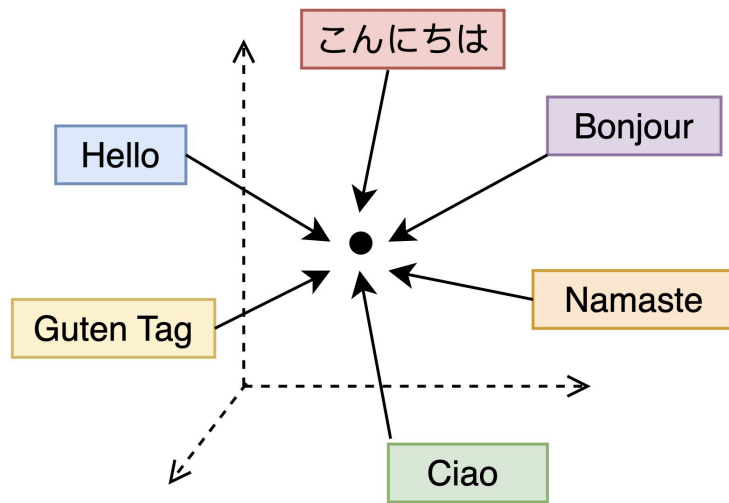
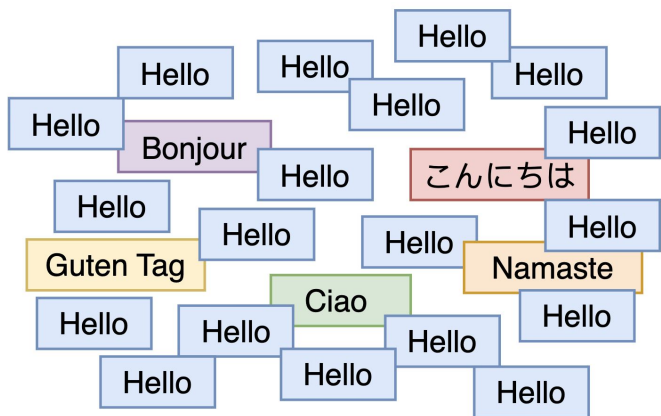
*\*Defined by various cultural psychologists at UPenn - come see my poster tomorrow!*

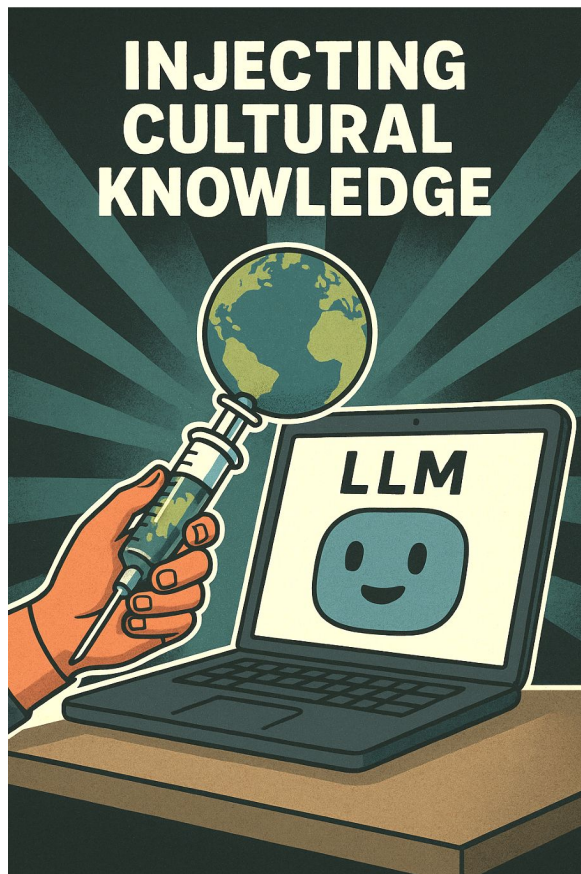
# LLMs are bad at incorporating cultural context



- **Considered: anxiety about unemployment**
- **Considered: future steps to getting a job**
- **Not considered: social roles & expectations of middle aged men in India**

# Datasets & training techniques cause LLMs to be Anglocentric





## One way to fix this: **Cultural Knowledge Injection**

1. Building a cultural norm corpus
2. Building a norm retrieval system
3. Norm incorporation
4. Human validation

# Building a Cultural Norm Corpus

# Utilizing large-scale datasets

Dataset	Overview
CultureBank [1]	33,000 cultural descriptors based on users' self-narratives on TikTok and Reddit.
Hollywood-Bollywood Shame & Pride [2]	10,000 contextualized movie conversations relating to pride and shame from underlying social expectations.
CulturalBench [3]	1,227 human-written and human-verified multiple choice questions for assessing LLMs' cultural knowledge.

[1] Yu Ying Chiu, Liwei Jiang, Bill Yuchen Lin, Chan Young Park, Shuyue Stella Li, Sahithya Ravi, Mehar Bhatia, Maria Antoniak, Yulia Tsvetkov, Vered Shwartz, and Yejin Choi. Culturalbench: a robust, diverse and challenging benchmark on measuring the (lack of) cultural knowledge of llms, 2024.

[2] Sunny Rai, Khushang Jilesh Zaveri, Shreya Havaladar, Soumna Nema, Lyle Ungar, and Sharath Chandra Guntuku. Social norms in cinema: A cross-cultural analysis of shame, pride and prejudice, 2025.

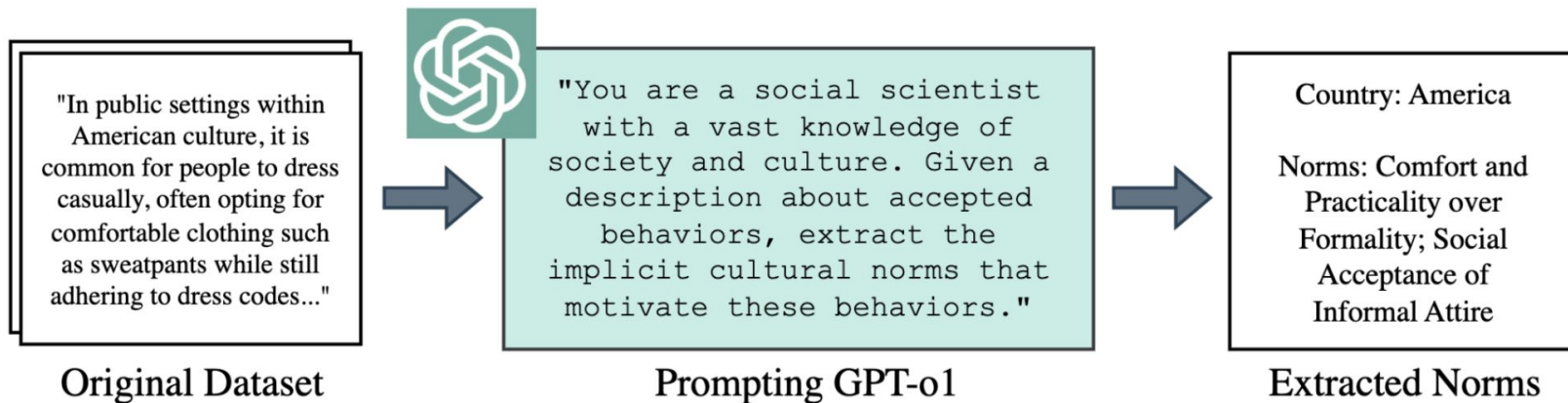
[3] Weiyan Shi, Ryan Li, Yutong Zhang, Caleb Ziems, Chunhua yu, Raya Horesh, Rogerio Abreu de Paula, and Diyi Yang. Culturebank: An online community-driven knowledge base towards culturally aware language technologies, 2024.



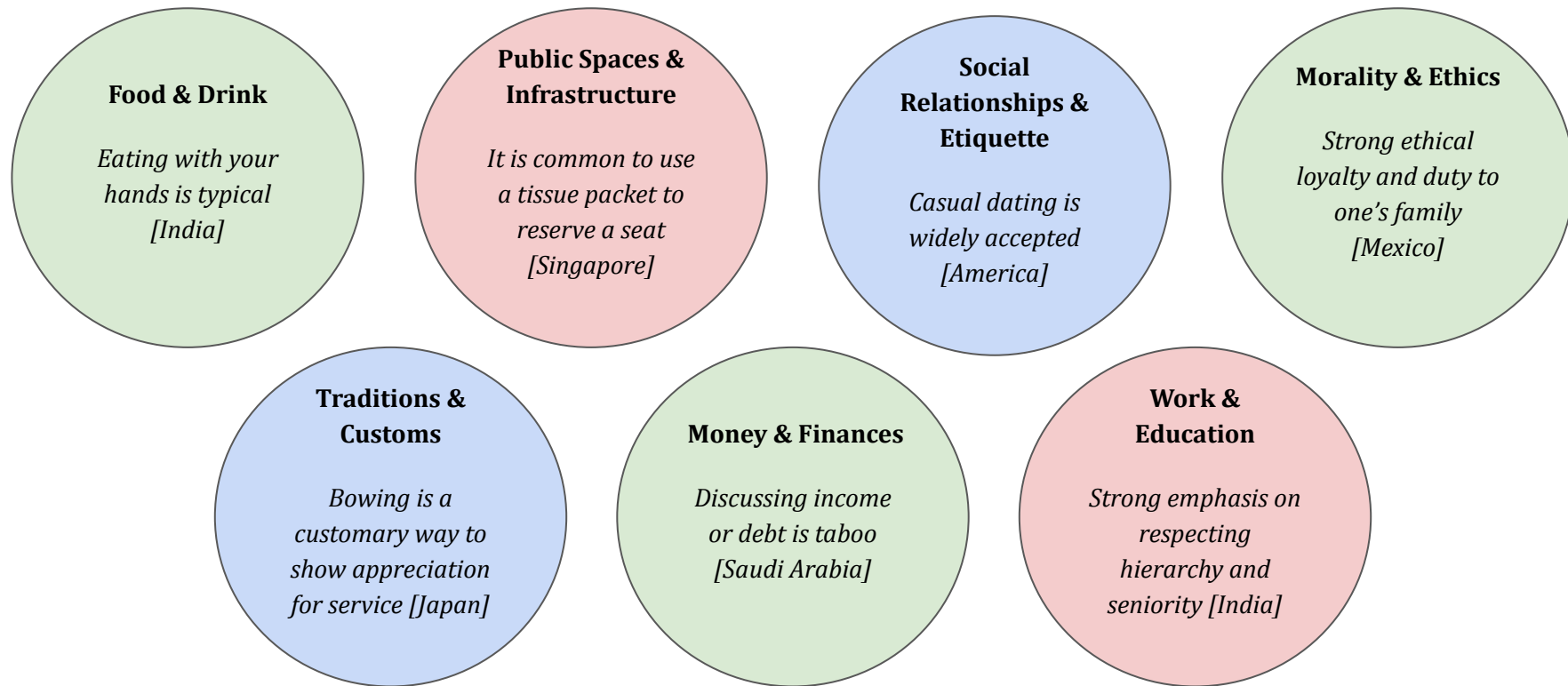
# Covering a variety of norms

Dataset	Norm Overview	Example
CultureBank [1]	First-person stories/posts based on cultural norms → represents <b>norms that confuse or surprise people</b>	<i>Bowing is a customary way to show appreciation for service [Japan]</i>
Hollywood-Bollywood Shame & Pride [2]	Extracted norms from movie subtitles → represents <b>norms that are pervasive and widely accepted</b>	<i>Men fulfilling the breadwinner role is a source of pride [India]</i>
CulturalBench [3]	Human written multiple choice questions designed to confuse LLMs → represents <b>norms that are tricky and not well-known</b>	<i>It is common to use a tissue packet to reserve a seat [Singapore]</i>

# Norm Extraction

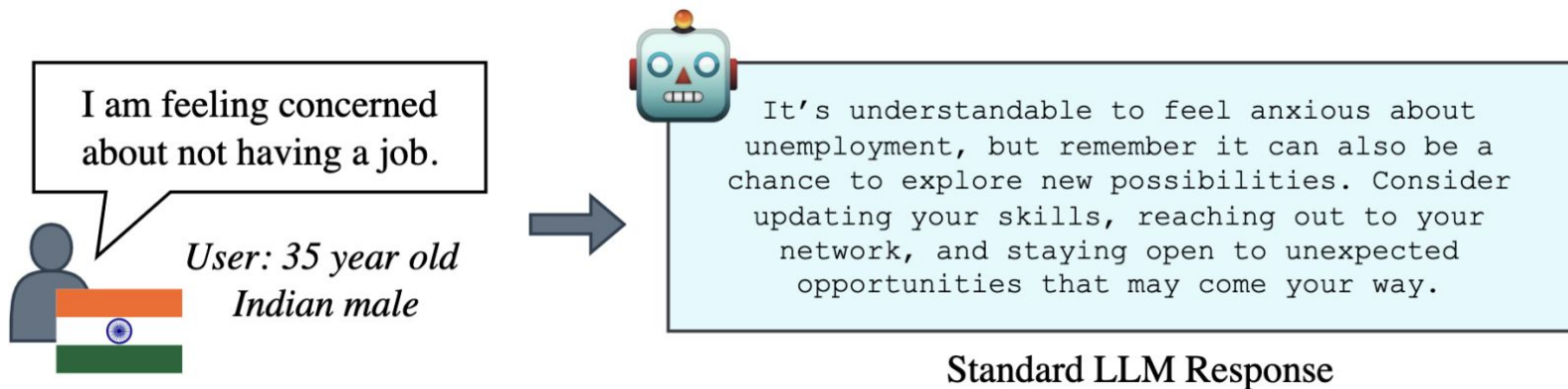


# Norm Embedding + Clustering



# Norm Retrieval & Incorporation

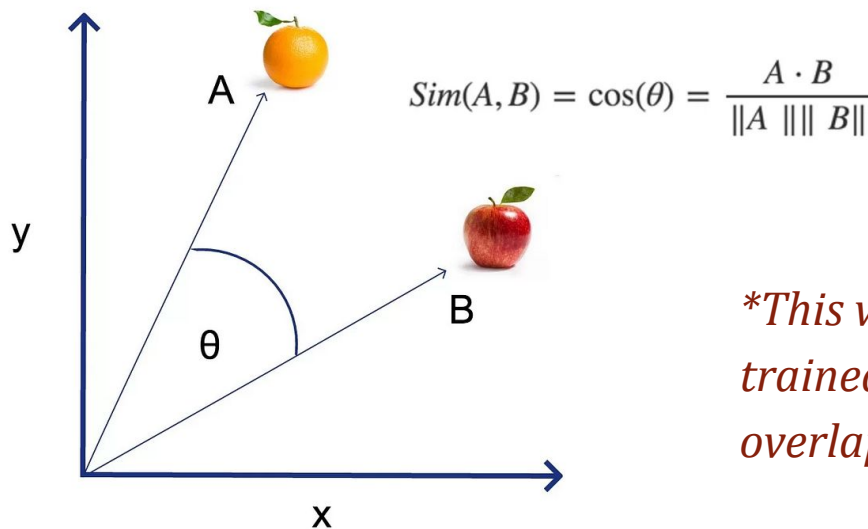
# LLMs are bad at incorporating cultural context



- **Considered: anxiety about unemployment**
- **Considered: future steps to getting a job**
- **Not considered: social roles & expectations of middle aged men in India**

# Why naive RAG works poorly...

Retrieval-Augmented Generation (RAG) works by embedding a query and selecting the item in the knowledge bank with the **highest cosine similarity to the query**.



*\*This works because embedding models are trained to make inputs with high semantic overlap be close in embedding space.*

# Why traditional RAG works poorly...

Retrieval-Augmented Generation (RAG) works by embedding a query and selecting the item in the knowledge bank with the **highest cosine similarity to the query**.

Q: What is the capital of France?

A: The capital of France is Paris

Typical applications of traditional RAG  
(information retrieval, scientific QA, etc.

→ ***queries & answers have high semantic overlap***

# Why traditional RAG works poorly...

Retrieval-Augmented Generation (RAG) works by embedding a query and selecting the item in the knowledge bank with the **highest cosine similarity to the query**.

Q: What is the capital of France?

A: The capital of France is Paris

Typical applications of traditional RAG  
(information retrieval, scientific QA, etc.)  
→ ***queries & answers have high semantic overlap***

Q: I am feeling concerned about not  
having a job.

A: Men fulfilling the breadwinner  
role is a source of pride.

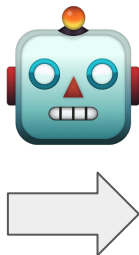
But... cultural norms ***don't usually have high semantic overlap*** with queries that require them.



# Solution: reverse engineer queries that enable RAG

*"Men fulfilling the breadwinner role is a source of pride"*

Cultural norm



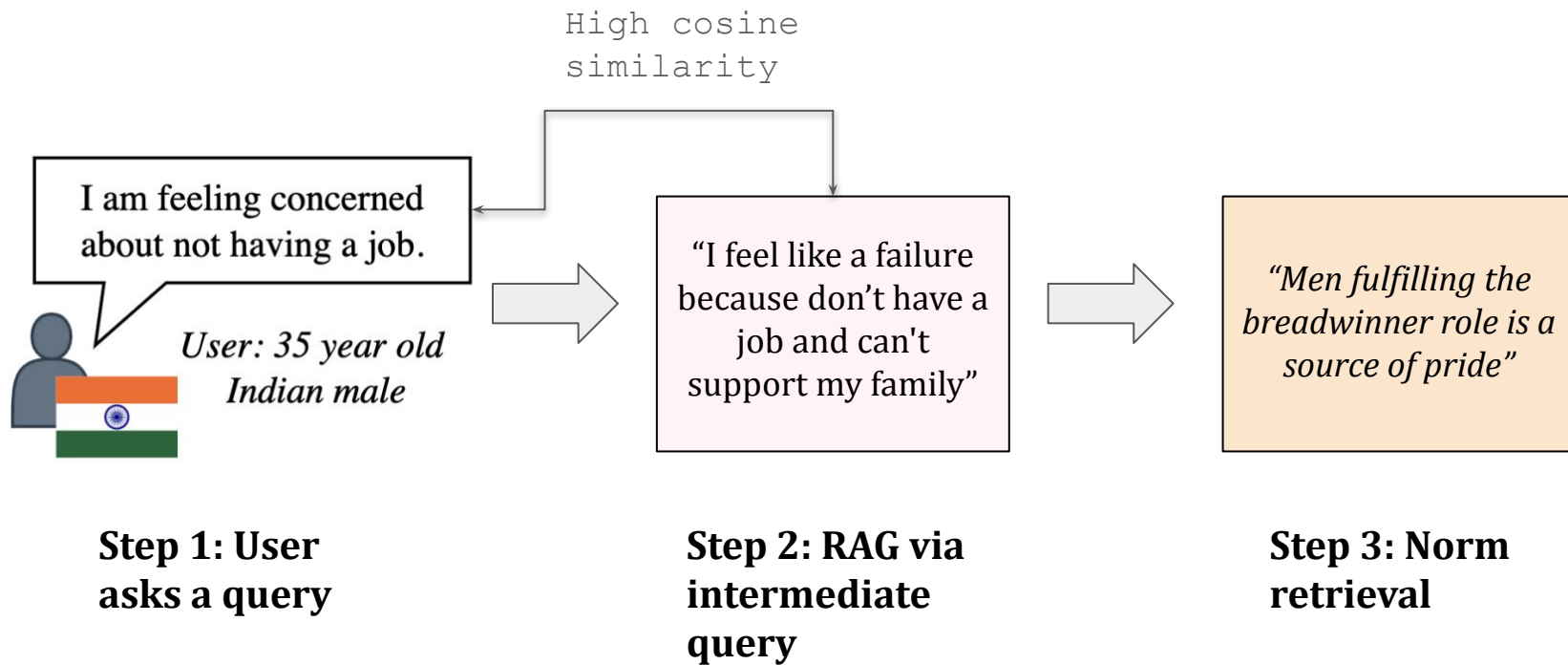
1. *I feel like a failure because don't have a job and can't support my family.*
2. *My wife earns more than me, and it makes me feel uncomfortable.*
3. *I'm the eldest son, but my younger sister just got a job and I didn't.*
- ...
10. *I'm considering becoming a stay-at-home dad, but I worry people will think I'm lazy.*

10 reverse engineered  
"intermediate" queries per norm

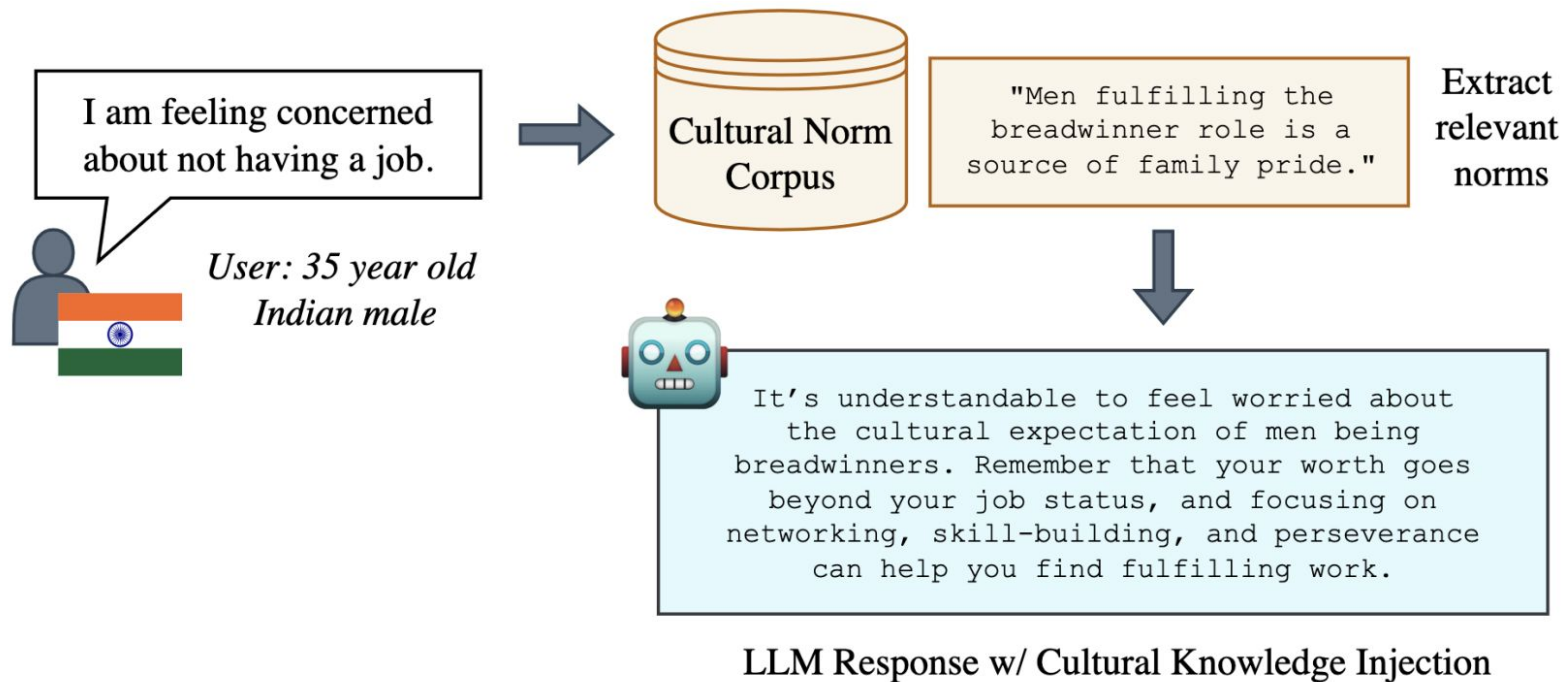
## Solution: **index norms by intermediate queries**

Intermediate Query	Embedding	Norm
<i>I feel like a failure because don't have a job and can't support my family.</i>	<0.03, 0.16, 0.12, 0.11 ...>	<i>“Men fulfilling the breadwinner role is a source of pride”</i>
<i>My wife earns more than me, and it makes me feel uncomfortable.</i>	<0.01, 0.20, 0.09, 0.15, ...>	
<i>My father worked two jobs to support us. I don't know if I'll ever live up to that.</i>	<0.02, 0.11, 0.07, 0.04 ...>	

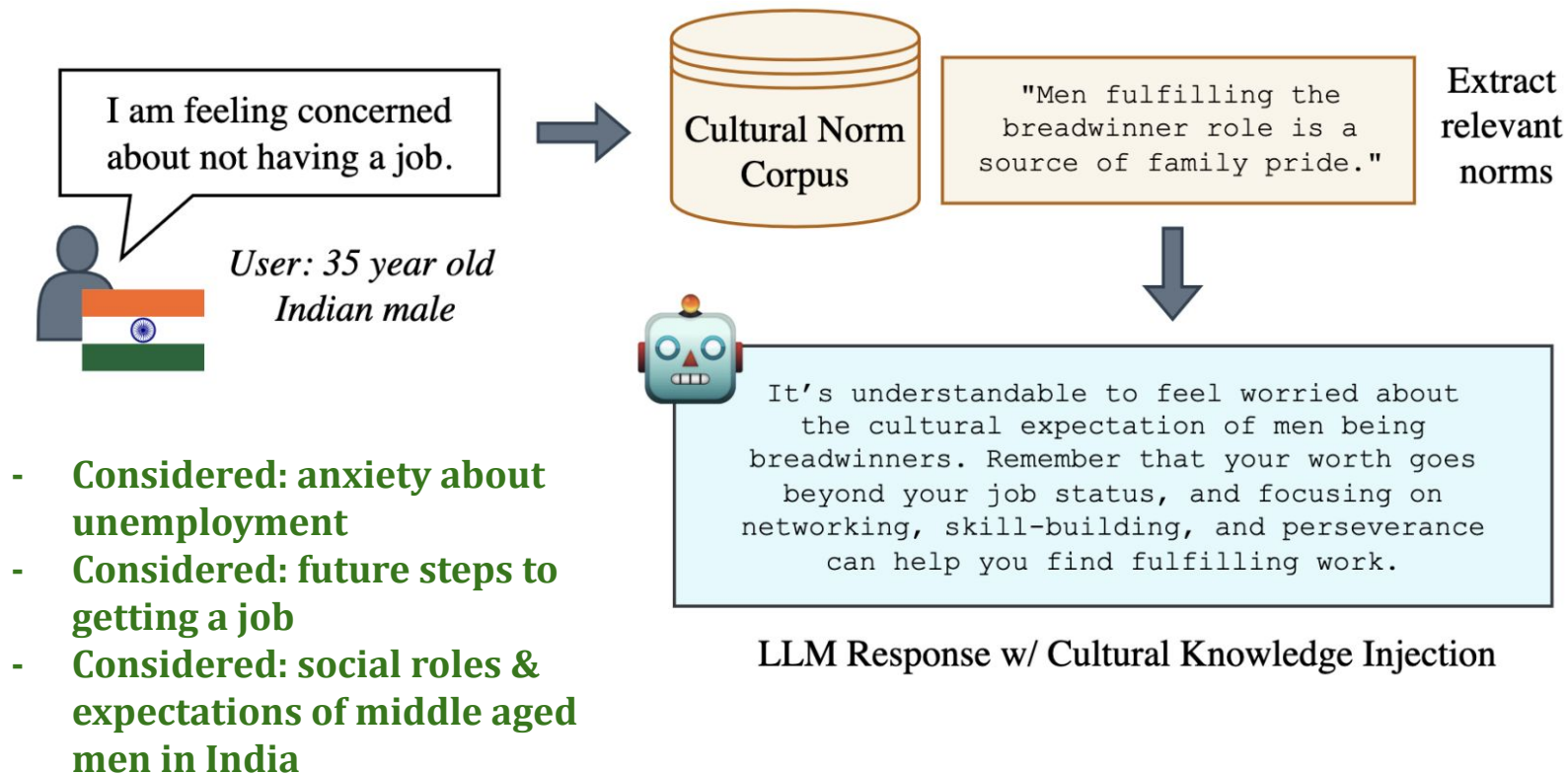
# Norm Retrieval



# Norm Incorporation



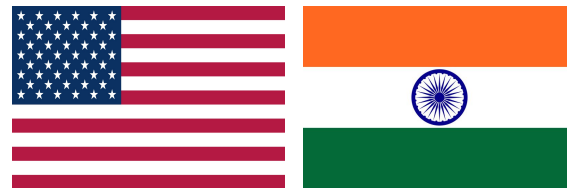
# Norm Incorporation



# Human Validation

# Small-scale Prolific study

- Ran Prolific study comparing advice from standard GPT-4 vs. knowledge-injected GPT-4
  - 100 random examples (*same per country*)
  - 2 annotators per example
- Indian annotators preferred injected responses  
71% of the time (Cohen's  $\kappa = 0.5$ ; moderate agreement)
- American annotators preferred injected responses  
65% of the time (Cohen's  $\kappa = 0.26$ ; fair agreement)

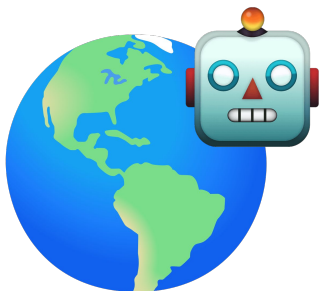


 **Prolific**

**Cultural knowledge injection  
can significantly improve text  
generations, particularly for  
non-Western countries like India!**

# Takeaways

- Current LLMs are super American-centric
- Applications of current LLMs require cultural sensitivity (*therapy, advice, education, and many more!*)
- Our contributions:
  - **Large-scale, fully automatic norm discovery from a variety of datasets**
  - **Fast, scalable system to retrieve & inject cultural knowledge into an LLM**



Future work: building LLM agents capable of culturally-aware conversations, validation on a larger sample, etc.





# Thank you!! Please ask questions



*My research group!*



**Find more info  
on my website!**



**Penn  
Engineering**  
UNIVERSITY of PENNSYLVANIA