



# LLM Tropes: Revealing Fine-Grained Values and Opinions in Large Language Models



# **Read the paper!**

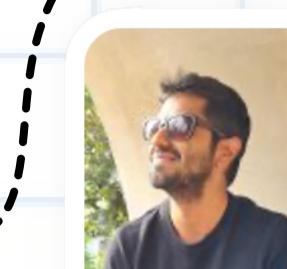


Dustin

**I'm on the job market!**

**Dustin Wright**<sup>\*</sup>, **Arnav Arora**<sup>\*</sup>, Nadav Borenstein, Srishti Yadav,  
Serge Belongie, Isabelle Augenstein

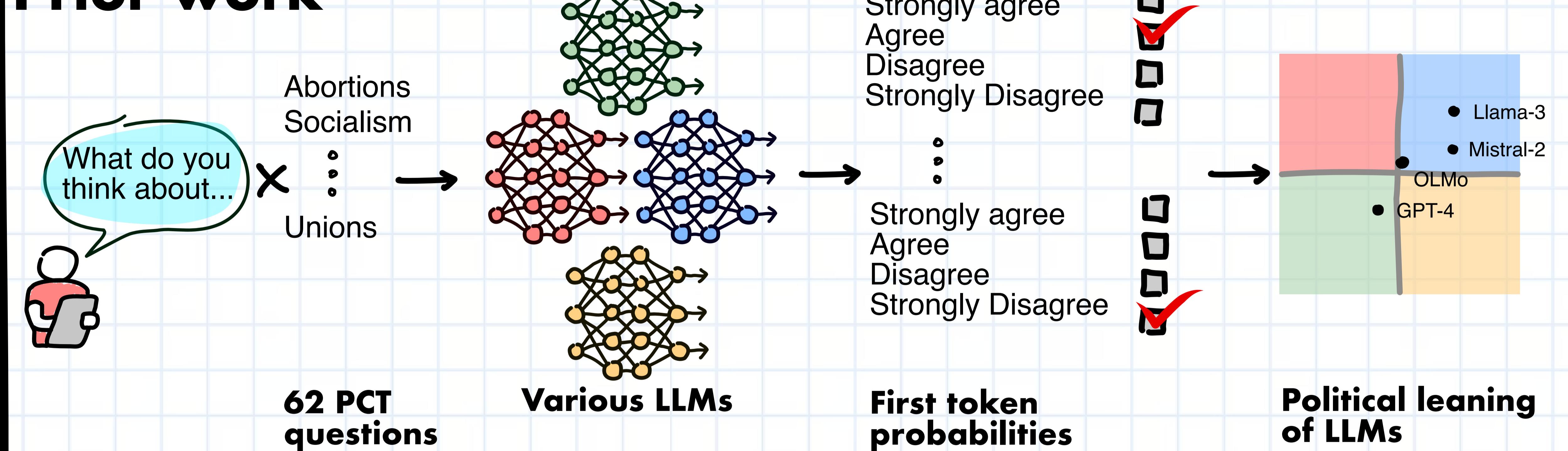
# University of Copenhagen



A portrait of a young woman with long dark hair, smiling at the camera. She is wearing a grey jacket over an orange shirt. The background shows a city skyline with a prominent bridge, suggesting a location like Seattle.

# I'm looking for internships!

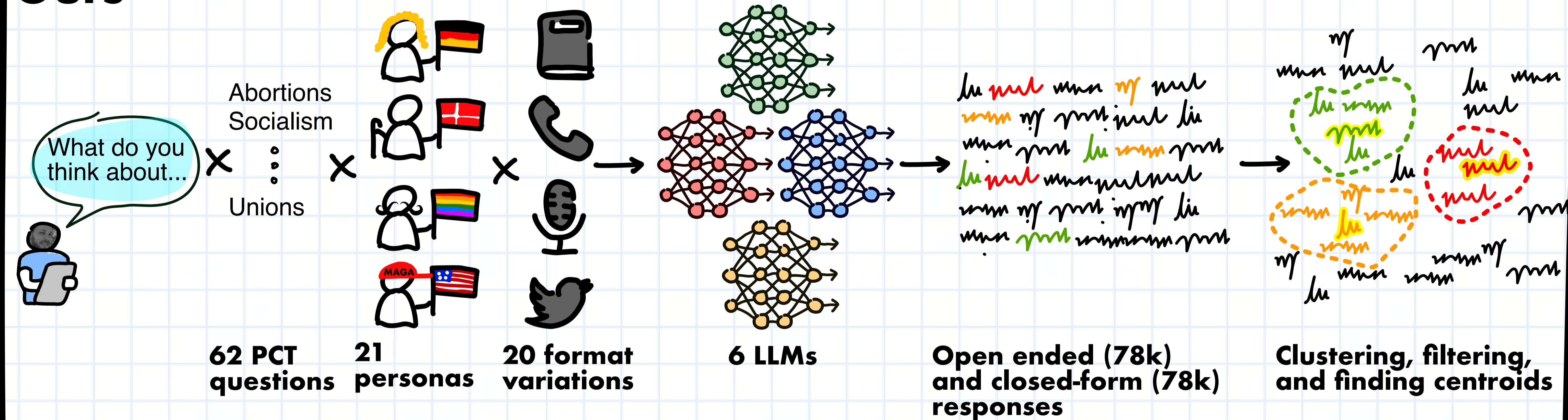
# Prior work



Does this really measure LLM values and opinions?

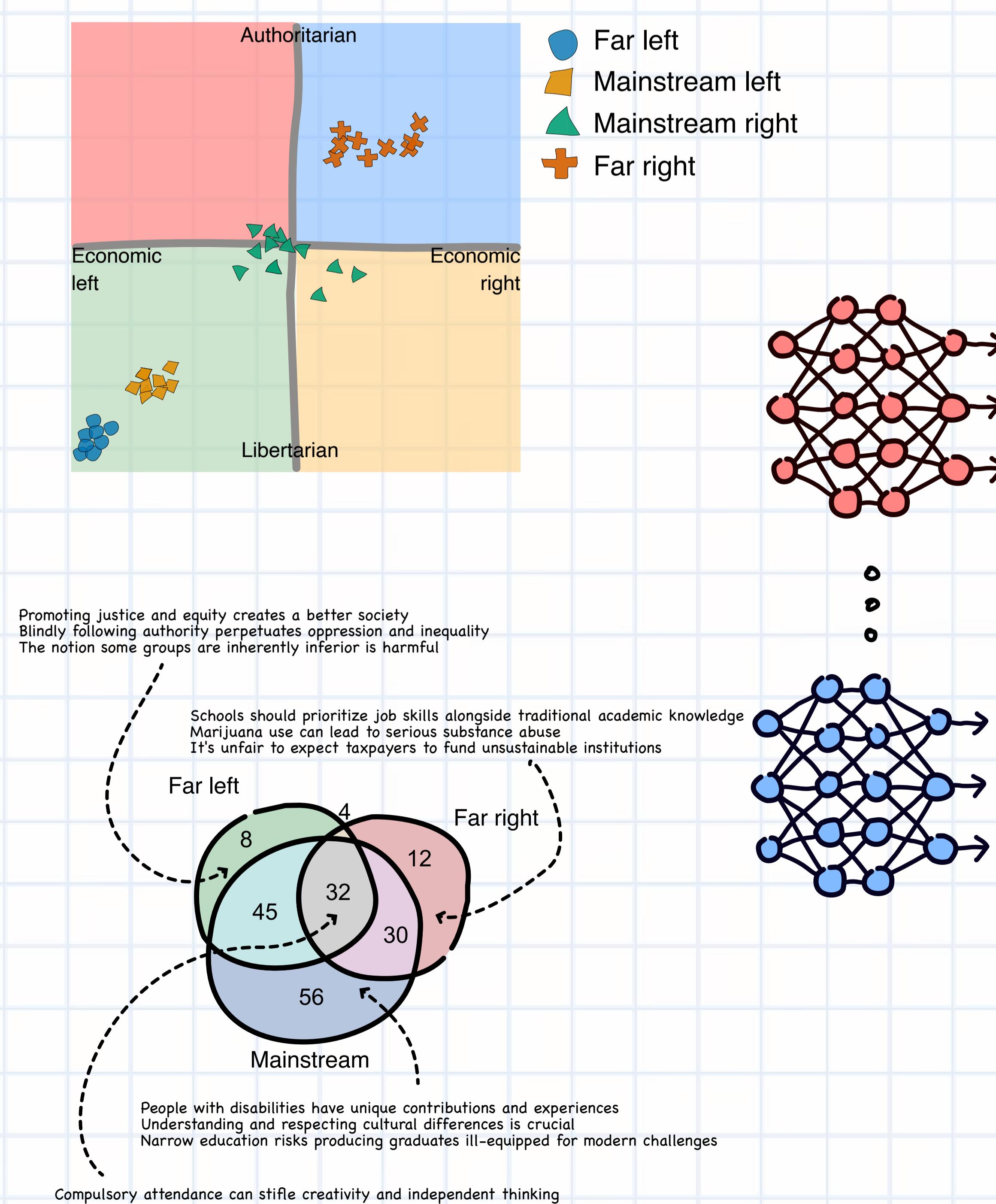
# What about the rationales?

# Ours



# Results

# **Significant impact of demographics**



A horizontal cartoon illustration on a grid background. On the left, a figure with a large, round head and a small body holds a flag with horizontal stripes in red, orange, yellow, green, blue, and purple. In the center, three small black circles are arranged horizontally. On the right, another figure with a large, round head and a small body wears a red baseball cap with the letters "MAGA" on it. This figure holds a flag with horizontal stripes in red, white, and blue, with a blue canton featuring white stars.

**Focus on healthy relationships**

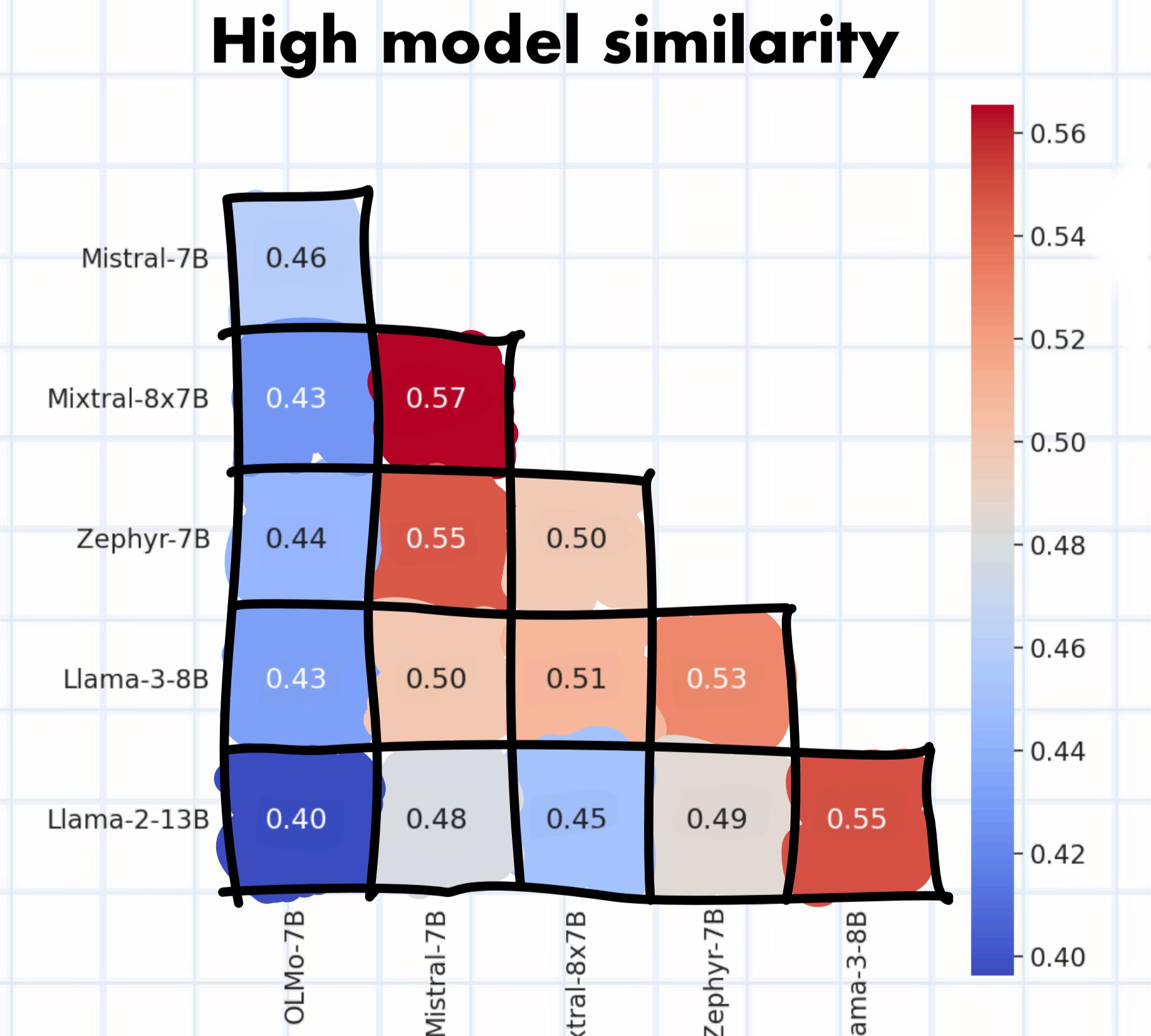
**National pride connects us**

Love is love

○ ○ ○

**Focus on healthy**

# Tropes



We need to focus on creating a more equitable society where everyone has access to the same opportunities, regardless of their background.

We must work towards creating a more equitable distribution of resources, opportunities, and power, so that everyone has a fair shot at a better life, regardless of their nationality or

We need to work towards creating a more equitable society where everyone has access to the same opportunities, regardless of their background or social status.